VII. Accounting for Uncertainty

The point estimates of price and purity reported in the main report were determined through a modeling process that is based on numerous assumptions that are not known with certainty. For example, the point estimates reflected in the main report may be biased because the STRIDE data may not be representative of the true distribution of price and purity. Thus, it is not known whether the model used here provides a good fit to the actual distribution of prices and purities in the country. All we can do is determine whether the model provides a relatively good fit of the distribution of these quantities observed in STRIDE, as was done in Appendices B and C. And, as is the case with any model, the model could lead to biased estimates if there are omitted variables or if there is model misspecification that is not easily identified from the STRIDE data. It is not possible for us to evaluate whether these sources of bias exist in our model estimates without using other data sources, a task that was outside the scope of the current project. However, it is strongly recommended that such an evaluation be conducted as a next step for future analyses using these data.

Even if the point estimates from the model could be shown to be asymptotically unbiased, a second source of uncertainty is the precision with which the coefficient estimates are determined, in light of the small samples. Every coefficient estimate has an associated standard error describing the precision with which it is being estimated, and this precision is not currently being accounted for in our estimates of the range of predicted prices. Estimation of this variability is a nontrivial task, given the complexity of the models being estimated, unless one is willing to impose strong distributional assumptions that cannot be empirically validated. There are computationally intensive methods, such as bootstrapping or Markov Chain Monte Carlo methods, that could be employed to construct more-precise intervals for the standardized predicted prices as well as the national index estimates themselves.[21] An example of these methods is presented below.

A third source of uncertainty is the estimate of the national price (purity) index. As discussed previously, the models identify independent variation in the distribution of prices (purities) within cities. This unique variation and the limited observation of prices (purities) across all cities in all periods imply that there is uncertainty in the national indices that are calculated. To try to represent some of the variation in the underlying distribution of weighted average indices, the main report presents estimates of the interquartile range of the city-specific prices (or purities) that constitute the national index. The interquartile range is used because it is easily constructed without assumptions regarding the underlying distribution of values and it is fairly easy to understand. The interquartile range simply reflects the highest and lowest values between which half of all the possible values fall. One-quarter of the remaining values are lower than the estimates provided by the interquartile range, and the other quarter are higher.

To generate estimates of the interquartile range, we randomly sample a city based on the city's population weights for a given year. For the sampled city, we take the standardized predictions

[21] For bootstrapping techniques, see: Efron, B., and R. Tibshirani (1993), "An Introduction to the Bootstrap," Chapman & Hall Ltd., London, New York. For Markov Chain Monte Carlo techniques, see Gelfand, A.E., and A.F.M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 , pp. 398–409.

for price and purity for the year. Based on sampling 5,000 city-standardized price and purity measures for each year for a given distribution level and drug, we identify the 25th, 50th (median), and 75th percentile price and purity observations. The 25th and 75th percentile observations provide the boundary of the interquartile range, which is reported for all the price and purity indices in the main report.

It should be emphasized that this interquartile range captures just a small part of the overall uncertainty in the estimates. It relies on the assumption that the price and purity measures for each city and each quarter and the model generating them are unbiased and precise. Future work should build upon this by estimating uncertainty in the price and purity values themselves, which means evaluating the other potential sources of uncertainty as well. To illustrate, we fit a Bayesian version of the hierarchical model using Markov Chain Monte Carlo (MCMC) to obtain the posterior distribution of all the model parameters as well as the national price (or purity) index based on very minimal assumptions of the prior distribution of these estimates. This approach involves first specifying a likelihood function for price (or purity), as was done before (Equations (1) through (5)) and then specifying prior distribution functions for all parameters in the model. The prior and the likelihood functions are multiplied to yield a posterior distribution for the model parameters; all statistical inferences are drawn from this posterior distribution. MCMC is a numerical integration technique for obtaining posterior distributions of interest by simulation. Bayesian hierarchical models were fit for the first quantity level for heroin and powder cocaine.[22] The models were estimated using WinBUGS Version 1.3 software to implement MCMC.[23] A proper but uninformative prior distribution was specified for all of the model parameters[24] and MCMC diagnostics were employed to ensure that the Markov chain had converged to the posterior distribution from its starting value.

Tables F.1 and F.2 in Appendix F compare, for the lowest quantity level for powder cocaine and heroin, the coefficient estimates resulting from the Bayesian hierarchical model with those obtained from a similar model estimated by the method used previously, restricted maximum likelihood (REML). These models differ from those discussed earlier in that they condition on year rather than on quarter, but other than that the likelihood functions are the same. The parameter estimates and their standard errors are remarkably similar across the two methods; the largest visible difference is for powder cocaine purity regression, but the estimates do not differ greatly from each other. The similarity in results despite fundamentally different modeling approaches is reassuring in light of the other problems associated with the data. National price indices were constructed using the predictions generated from the Bayesian hierarchical model data, and the interquartile range for the posterior distribution of the national price index was calculated, as well as the 95 percent posterior probability interval. Figures F.1 and F.2 in

---

[22] At this point, the MCMC for the price models was implemented with predicted purity constructed using the standard techniques in this report instead of the MCMC approach. In addition, this was done with the final sample after having deleted extreme residuals rather than deleting extreme residuals with the MCMC technique.
[23] Spiegelhalter, D., A. Thomas, and N. Best (2000), *WinBUGS Version 1.3 User Manual*, MRC Biostatistics Unit, Cambridge, UK.
[24] Regression parameters had normal prior distributions centered at 0 in all models, with prior variances of 100 in the purity model and 10 in the price model (the difference is due to the fact that purity ranges from 0 to 100, while price is modeled on the log scale); precision (reciprocal of the variance) parameters were assumed to follow Gamma distributions with mean 1 and variances 10 in the price model (e.g., Gamma(0.1,0.1) and 100 in the purity model (e.g., Gamma(0.01,0.01)).

Appendix F show the interquartile range of the posterior distribution, indicated by the 25 percent and 75 percent bands in the figures. These ranges are substantially smaller than those reported using the non-Bayesian methods. The trends in the national price estimates derived from this model, however, are again consistent with those reported previously. The 95 percent posterior probability interval for the national price index is also shown on each figure and is indicated by the lower and upper bounds at 2.5 percent and 97.5 percent, respectively. The advantage of using this type of method to compute an uncertainty estimate for the national price index is that the resulting posterior distributions of the national price indices could be used to test whether the national price index significantly differs at any two given time points.

This estimation merely serves as an example of the type of analysis that could be done to identify some of the uncertainty in the current method if greater time and resources were available. The fact that the model coefficients and predicted trend lines were consistent with what was reported previously is reassuring, but future work should be done to determine whether similar findings hold for the other substances and to test some of the remaining assumptions in the model. For example, the current price model includes a single estimate of expected purity and ignores any uncertainty that might exist with this single estimate. In future work, it would be interesting to test the sensitivity of the results to this assumption by estimating a joint model of price and purity that allows for uncertainty in expected purity.