Careful with that prior

Suppose we have a data generated by the linear guassian Bayesian network below. The goal is to estimate the causal effect of X on Y, that is identifying the value of the coefficient w_{yx}. The variable U is latent.

$$\mathrm{P}(U) = N(0, v_u) \tag{6.1}$$
$$\mathrm{P}(Z|U) = N(w_{zu}U, v_z) \tag{6.2}$$
$$\mathrm{P}(X|Z) = N(w_{xz}Z, v_x) \tag{6.3}$$
$$\mathrm{P}(Y|U, Z, X) = N(w_{yu}U + w_{yz}Z + w_{yx}X, v_y) \tag{6.4}$$

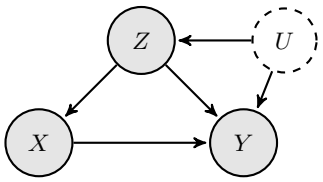

Figure 6.1

As each variable is a linear function of its parents, with gaussian noise, resulting in a joint distribution $\mathrm{P}(U, Z, X, Y)$ that is multivariate normal. Marginalising out U and conditioning on X and Z yields,

$$Y \sim N(w_{yx}X + \beta Z, \varepsilon) \tag{6.5}$$

where

$$\beta = w_{yz} + \frac{w_{yu}w_{zu}}{w_{zu}^2 + \frac{v_z}{v_u}}, \text{ and} \tag{6.6}$$
$$\varepsilon = v_y + w_{yu}^2 v_u - \frac{v_u^2 w_{zu}^2 w_{yu}^2}{v_z^2 + v_u^2 w_{zu}^2} \tag{6.7}$$

The causal effect of X on Y is identifiable (even without the linear gaussian assumptions) as Z satisfies the backdoor criterion and the expectation of the coefficient for X in a standard OLS regression of Y against X and Z is w_{yx}. However, the causal effect of Z on Y is not identifiable due the presense of the unobserved confounder U. The coefficient β captures both the causal relationship between Z and Y, w_{yz} and the indirect relationship through U, which we cannot seperate without observing U.

If we do a Bayesian regression of X and Z against Y and use intuition based on the causal relationship between Z and Y to select a prior for β we are likely to select something centered around w_{yz}, but this is not correct and can introduce bias into our estimate of the coefficent for w_{yx}. The prior for β should be based on our belief about how we expect Z to be *associated* with Y after marginalising out any confounding variables.

This result is counter intuative. When fitting a model to estimate the affect of marketing on sales, it is natural to put a prior on the coefficient for price that is strongly negative because we believe increasing price decreases sales. However, our prior should be on the expected association between price and sales, holding fixed all the other variables in the model, not the causal relationship between price and sales. An alternative is to explicitly include U as a latent variable in the model. However, this comes at a computational cost with little benifit for estimation, unless we have good knoweledge of the relationships between U, Z and Y. More fundamentally, at whatever level we decide to stop adding latent variables and specify a prior, we can introduce this form of bias if priors are selected naivily on the basis of causal intuition.

6.1 Example

$$U \sim N(0, 1) \tag{6.8}$$
$$Z \sim N(2U, 0.09) \tag{6.9}$$
$$X \sim N(0.5Z, 1) \tag{6.10}$$
$$Y \sim N(3U + -Z + 0.5X, 0.25) \tag{6.11}$$

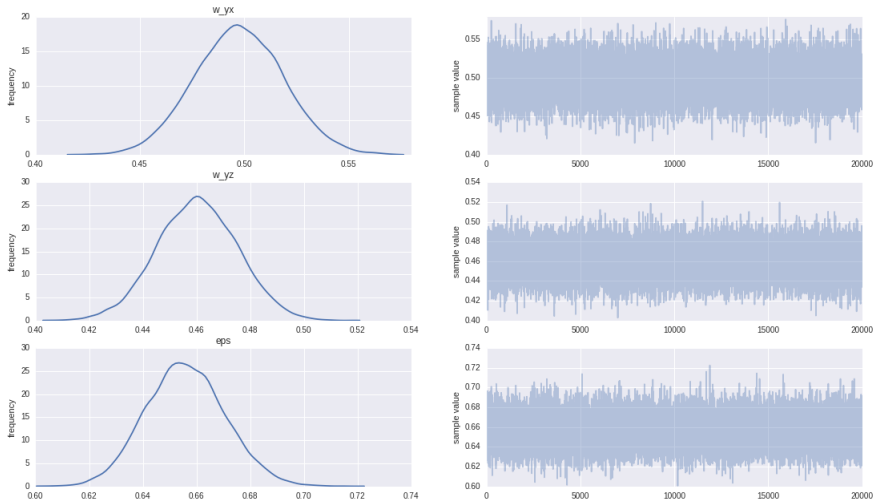

Figure 6.2: no prior, posterior on w_{yx} centered around w_{yx}

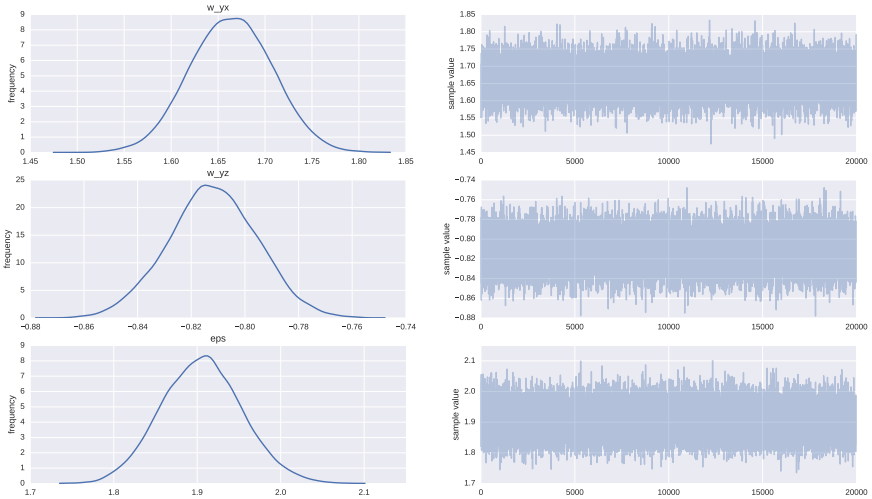

Figure 6.3: Prior around w_{yz}, $\beta \sim N(w_{yz}, \sigma = 0.5)$. The posterior on w_{yx} is biased away from its true value