Wald Test for Common Mean Probabilities across Geographic Areas

Michael LaForest Steven Stern

September 2016

1 Introduction

There is a large and growing literature on "small area estimation." Applications include, for example, Ghosh and Rao (1994), Rao (1999), Alderman et al. (2002), Elbers, Lanjouw, and Lanjouw (2003), Elbers et al. (2007), and Tarozzi and Deaton (2009) in development economics; Malec et al. (1997), Leroux, Lei, and Breslow (1999), Twigg, Moon, and Jones (2000), Legler et al. (2002), Mendez-Luck et al. (2007), Choy et al. (2008), and Congdon (2009) in health; Jarjoura et al. (1993), Kessler et al. (1998), Messer et al. (2004), Congdon (2006), Ellis et al. (2009); Konrad et al. (2009), Stern (2014), and Johnson et al. (2016) in mental health; and Citro and Kalton (2000) in local allocation of resources. However, the outcome of interest may be discrete, and there may be important area-specific effects affecting the precision of both the small area estimates and any test of meaningful variation across small areas due to observed covariates. If the covariates do not explain any variation in outcomes across small areas, then the exercise is not useful. While many papers model random effects when individual location data are available (e.g., Biggeri et al., 1999; Ferrándiz, López, and Sanmartin, 1999; Leroux, Lei, and Breslow, 1999; Prasad and Rao, 1999; Banerjee, Wall, and Carlin, 2003; MacNab, 2004; Malec and Müller, 2008), much of the literature does not allow for any small area-specific effects (e.g., Congdon, 2006; Kessler et al., 1998; Hudson, 2009; Konrad, 2009). However, Tarozzi and Deaton (2009) argue convincingly for the inclusion of small area-specific effects. If such effects do exist and they are ignored, then standard errors of parameter estimates are biased downward and, in nonlinear models, parameter estimates are inconsistent. Following Elbers, Lanjouw, and Lanjouw (2003), Stern (2014) suggests how to estimate such effects using sample weighting information when no spatial information is provided.[1]

Much of the literature focuses on continuous outcome measures such as income (e.g., Malec et al.,1997; Leroux, Lei, and Breslow, 1999; Banerjee, Wall, and Carlin, 2003; Elbers, Lanjouw, and Lanjouw, 2003; Malec and Müller, 2008;

[1] This idea was suggested to us by Wayne-Roy Gayle. Malec and Müller (2008) use a similar methodology but with restricted information on geography in the NHIS.

Opsomer et al., 2008), but there is a growing literature where the outcome is a qualitative or discrete variable (see, for example, Ghosh et al., 1998; MacNab, 2004; Stern, 2014; Johnson et al., 2016). The interest in qualitative or discrete outcomes adds some extra complications because of the nonlinear way in which both observed covariates and random effects affect outcomes.

Most of the literature avoids the question of testing for meaningful variation due to observed covariates across small areas. Much of the literature uses Bayesian methods (see, for example, Ghosh and Rao,1994; Malec et al.,1997; Ghosh et al.,1998; Leroux, Lei, and Breslow, 1999; Prasad and Rao, 1999; Banerjee, Wall, and Carlin, 2003; MacNab, 2004; Malec and Müller, 2008), and testing is not a natural concept in Bayesian analysis. But Jarjoura et al. (1993), Elbers, Lanjouw, and Lanjouw (2003), Konrad et al. (2009), Stern (2014), and Johnson et al. (2106a) all estimate their model parameters using classical methods; of those, only Tarozzi and Deaton (2009), Stern (2014), and Johnson et al. (2106a) allow for random effects, and only Johnson et al. (2106a) test for significant variation across small areas.

In this paper, we show how one can construct a Wald test in the presence of small area-specific randomness also taking into account the likely nonlinear structure of the relevant probability model. The methodology is applied to estimates of mental illness prevalence for small areas in Virginia described in more detail in Johnson et al. (2016a).

The paper proceeds as follows. In Section 2, we use a logit model to construct the distribution of the mean sample probability for each of J subgroups of a sample population. In Section 3, we conduct the same analysis when each subgroup has its own independent random effect. In Section 4, we conduct the same analysis once again when there is correlation between the random effects across groups. Section 5 describes how to use this distribution to conduct a Wald test to determine whether variation in mean sample probabilities across groups is statistically significant in the presence of correlated random effects. Section 6 discusses an empirical example of how this methodology can be applied, and Section 7 discusses how it might (or might not) be applied in other papers in the literature.

2 Analysis Without Small Area-Specific Random Effect

Consider a probit model,

$$\begin{aligned} y_i &= 1\left(x_i\beta + u_i > 0\right), \\ u_i &\sim iidN\left(0,1\right), \\ i &= 1,2,..,\widehat{n}. \end{aligned}$$

Let $\widehat{\beta}$ be the probit estimator of β. We know that

$$\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)\sim N\left(0,\Omega_\beta\right)$$

for known Ω_β (Maddala, 1983).

Next, divide up the population J into groups, indexed by j, and define J_j as the subset of the population in group j with

$$\bigcap_j J_j=\varnothing \text{ and } \bigcup_j J_j=J.$$

Similarly, define $\widehat{J}_j$ as the subset of the sample in group j. The mean population probability and mean sample probability for subset j are

$$\begin{aligned} p_j &= \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta\right), \qquad (1)\\ \widehat{p}_j &= \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\widehat{\beta}\right), \end{aligned}$$

respectively. Define $p=(p_1,p_2,..,p_N)'$ and $\widehat{p}$ analogously. Decompose $\widehat{p}_j-p_j$ as

$$\begin{aligned} & \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\widehat{\beta}\right)-\frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta\right) \qquad (2)\\ =\ & \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\left(\Phi\left(x_i\widehat{\beta}\right)-\Phi\left(x_i\beta\right)\right)+\frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\beta\right)-\frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta\right). \end{aligned}$$

The first term on the right is the deviation associated with having just an estimate of β averaged over subsample $\widehat{J}_j$, and the second term is the deviation due to sampling conditional on the true value of β. Taking a first order Taylor series approximation, equation (2) can be written as

$$\frac{1}{\widehat{n}_j\sqrt{\widehat{n}}}\sum_{i\in \widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)+e_j \qquad (3)$$

where

$$\begin{aligned} Ee_j &= 0,\\ Var\left(e_j\right) &= \frac{1}{n_j^2}\sum_{i\in J_j}\Phi\left(x_i\beta\right)\left(1-\Phi\left(x_i\beta\right)\right). \end{aligned}$$

Then

$$\sqrt{\widehat{n}}\left(\widehat{p}-p\right)\sim N\left(0,\Omega_p\right)$$

where

$$
\begin{aligned}
\Omega_{pjk} &= \widehat{n} E\left[\frac{1}{\widehat{n}_j\sqrt{\widehat{n}}}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)+e_j\right]\cdot \\
&\quad \left[\frac{1}{\widehat{n}_k\sqrt{\widehat{n}}}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)+e_k\right]'
\end{aligned}
$$

$$
= E\left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)+e_j\right]\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)+e_k\right]'
$$

$$
= E\left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right]+Ee_je_k'
$$

$$
\begin{aligned}
&= \left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right] \\
&\quad +1\left(j=k\right)\frac{1}{n_j^2}\sum_{i\in J_j}\Phi\left(x_i\beta\right)\left(1-\Phi\left(x_i\beta\right)\right) \\
&\approx \left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right].
\end{aligned}
$$

Our goal is to test $H_0: p_1 = p_2 = \cdots = p_J$ vs. $H_0: p_1 \neq p_2 \neq \cdots \neq p_J$; we refer to this null hypothesis as H_0^*.

3 Analysis With Area-Specific Independent Random Effect

Now we add a j-specific random effect. Define $j(i)$ as the subset to which i belongs. Then, we can write our random-effects probit model as

$$
\begin{aligned}
y_i &= 1\left(x_i\beta+\sigma_w w_{j(i)}+u_i>0\right), \\
w_j &\sim iidN\left(0,1\right), \\
u_i &\sim iidN\left(0,1\right).
\end{aligned}
$$

Let $\widehat{\beta}$ be the random-effects probit estimator of β. We know that

$$
\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)\sim N\left(0,\Omega_\beta\right)
$$

for known Ω_β. Analogous to equation (1), the mean population probability and mean sample probability for subset j are

$$p_j = \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta+\sigma_w w_{j(i)}\right), \tag{4}$$
$$\widehat{p}_j = \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\widehat{\beta}\right),$$

respectively. The relevant decomposition of $\widehat{p}_j - p_j$ involves some extra terms because of the inclusion of the random effect. In particular, write $\widehat{p}_j - p_j$ as

$$\begin{aligned} & \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\widehat{\beta}\right) - \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta+\sigma_w w_j\right) \qquad (5)\\ = & \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\left(\Phi\left(x_i\widehat{\beta}\right)-\Phi\left(x_i\beta\right)\right) + \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\left(\Phi\left(x_i\beta\right)-\Phi\left(x_i\beta+\sigma_w w_j\right)\right) \\ & +\frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\beta+\sigma_w w_j\right) - \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta+\sigma_w w_j\right). \end{aligned}$$

The first term on the right is the deviation due to having only an estimate of β evaluated at $\sigma_w = 0$, the second term is the deviation due to $\sigma_w > 0$ evaluated at the true value of β, and the last two terms together form the deviation to due random sampling evaluated at the true value of (β, σ_w). Analogously to equation (3), taking a first-order Taylor series approximation with respect to $\left(\widehat{\beta}, \sigma_w\right)$ evaluated at $(\beta, 0)$, equation (5) can be written as

$$= \frac{1}{\widehat{n}_j\sqrt{\widehat{n}}}\sum_{i\in \widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right) - \frac{\sigma_w w_j}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\phi\left(x_i\beta\right) + e_j \tag{6}$$

where

$$\begin{aligned} e_j &= \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\beta+\sigma_w w_j\right) - \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta+\sigma_w w_j\right) \\ &= \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\beta\right) + \frac{\sigma_w w_j}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\phi\left(x_i\beta\right) - \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta\right) - \frac{\sigma_w w_j}{n_j}\sum_{i\in J_j}\phi\left(x_i\beta\right) \\ &= \frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\Phi\left(x_i\beta\right) - \frac{1}{n_j}\sum_{i\in J_j}\Phi\left(x_i\beta\right) + \sigma_w w_j\left(\frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\phi\left(x_i\beta\right) - \frac{1}{n_j}\sum_{i\in J_j}\phi\left(x_i\beta\right)\right). \end{aligned}$$

Note that, in the last term, $\sigma_w w_j$ is $O(1)$, and

$$\frac{1}{\widehat{n}_j}\sum_{i\in \widehat{J}_j}\phi\left(x_i\beta\right) - \frac{1}{n_j}\sum_{i\in J_j}\phi\left(x_i\beta\right)$$

is $O\left(1/\sqrt{\widehat{n}_j}\right)$. Thus, as long as $\widehat{n}_j \to \infty$ as $\widehat{n} \to \infty$,

$$\begin{aligned} Var\left(e_j\right) &= \frac{1}{\widehat{n}_j^2} \sum_{i \in \widehat{J}_j} \Phi\left(x_i \beta\right)\left(1 - \Phi\left(x_i \beta\right)\right) + \sigma_w^2 O\left(\frac{1}{\widehat{n}_j^2}\right) \\ &\approx \frac{1}{\widehat{n}_j^2} \sum_{i \in \widehat{J}_j} \Phi\left(x_i \beta\right)\left(1 - \Phi\left(x_i \beta\right)\right). \end{aligned}$$

Then, the asymptotic distribution of $\widehat{p}$ is

$$\sqrt{\widehat{n}}\left(\widehat{p} - p\right) \sim N\left(0, \Omega_p\right)$$

with

$$\begin{aligned} \Omega_{pjk} &= \widehat{n} E\left[\frac{1}{\widehat{n}_j \sqrt{\widehat{n}}} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) x_i' \sqrt{\widehat{n}}\left(\widehat{\beta} - \beta\right) - \frac{\sigma_w w_j}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) + e_j\right] \cdot \\ &\quad \left[\frac{1}{\widehat{n}_k \sqrt{\widehat{n}}} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) x_i' \sqrt{\widehat{n}}\left(\widehat{\beta} - \beta\right) - \frac{\sigma_w w_k}{\widehat{n}_j} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right) + e_k\right]' \\ &= E\left[\frac{1}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) x_i' \sqrt{\widehat{n}}\left(\widehat{\beta} - \beta\right) - \frac{\sigma_w w_j}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) + e_j\right] \cdot \\ &\quad \left[\frac{1}{\widehat{n}_k} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right) x_i' \sqrt{\widehat{n}}\left(\widehat{\beta} - \beta\right) - \frac{\sigma_w w_k}{\widehat{n}_j} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right) + e_k\right]' \\ &= E\left[\frac{1}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) x_i'\right] \Omega_\beta \left[\frac{1}{\widehat{n}_k} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right) x_i\right] \\ &\quad + \sigma_w^2 \left(\frac{1}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right)\right)\left(\frac{1}{\widehat{n}_k} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right)\right) + E e_j e_k' \\ &= \left[\frac{1}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right) x_i'\right] \Omega_\beta \left[\frac{1}{\widehat{n}_k} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right) x_i\right] \\ &\quad + \sigma_w^2 \left(\frac{1}{\widehat{n}_j} \sum_{i \in \widehat{J}_j} \phi\left(x_i \beta\right)\right)\left(\frac{1}{\widehat{n}_k} \sum_{i \in \widehat{J}_k} \phi\left(x_i \beta\right)\right) \\ &\quad + 1\left(j = k\right) \frac{1}{n_j^2} \sum_{i \in J_j} \Phi\left(x_i \beta\right)\left(1 - \Phi\left(x_i \beta\right)\right) \end{aligned}$$

$$
\begin{aligned}
\approx \quad & \left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right] \\
& +\sigma_w^2\left(\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)\right)\left(\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)\right).
\end{aligned}
$$

4 Analysis With Area-Specific Dependent Random Effect

Now we allow the j-specific random effect in the model to exhibit spatial correlation. Define $w = (w_1, w_2, .., w_N)'$. We can write our random-effects probit model as

$$
\begin{aligned}
y_i &= 1\left(x_i\beta + \sigma_w w_{j(i)} + u_i > 0\right), \\
u_i &\sim iidN\left(0,1\right), \\
w &\sim N\left(0,\Omega_w\right).
\end{aligned} \tag{7}
$$

Assume that

$$\Omega_{wjk} = \rho\left(d_{jk}\right)$$

where d_{jk} is the spatial distance between subsets j and k and $\rho(\cdot)$ is the function presented[2] in Figure 1 and estimated in Stern et al. (2010) with $\rho(0) = 1$, $\rho'(d_{jk}) \leq 0$.[3] In some literature, this specification is called isotropic.[4] Other papers, especially those using Bayesian methods, also model spatial correlation (e.g., Ghosh, et al.,1998; Ferrándiz, López, and Sanmartin, 1999; Leroux, Lei, and Breslow,1999; Banerjee, Wall, and Carlin, 2003; MacNab, 2004) though in ways somewhat different than ours and requiring more information than available to us.[5] Let $\widehat{\beta}$ be the random-effects probit estimator of β under the assumption that $\rho(d_{jk}) = \rho_L(d_{jk})$ where $\rho_L(d_{jk}) = 0$ for all $d_{jk} > 0$. We know that

$$\sqrt{n}\left(\widehat{\beta} - \beta^*\right) \sim N\left(0,\Omega_\beta\right)$$

for known Ω_β where

$$\beta^* = plim\widehat{\beta} \neq \beta.$$

However, as $\rho(\cdot) \to \rho_L(\cdot)$, $\beta^* \to \beta$. We assume that, since $\sup_{d>0}\rho(d)$ is relatively small (0.2), β^* is close to β. The mean population probability and

[2] The figure is taken straight from page 50 of Stern et al. (2010).

[3] Note that the estimate of $\rho(0)$ is approximately 0.2. This is a feature of the kernel estimation procedure. The correct interpretation is that $\rho(0) = 0$ and $\rho(\cdot)$ decreases quickly for small values of its argument.

[4] See Cressie (1993) for a generalization in multiple dimensions.

[5] In the ecology literature, there is a tendency to model spatial dependence using polynomial functions of longitude and latitude (e.g., Legendre and Legendre, 1998; Lichstein et al., 2002; Rossi et al., 1992). This approach imposes much more structure than our approach, while ours requires spatial stationarity.

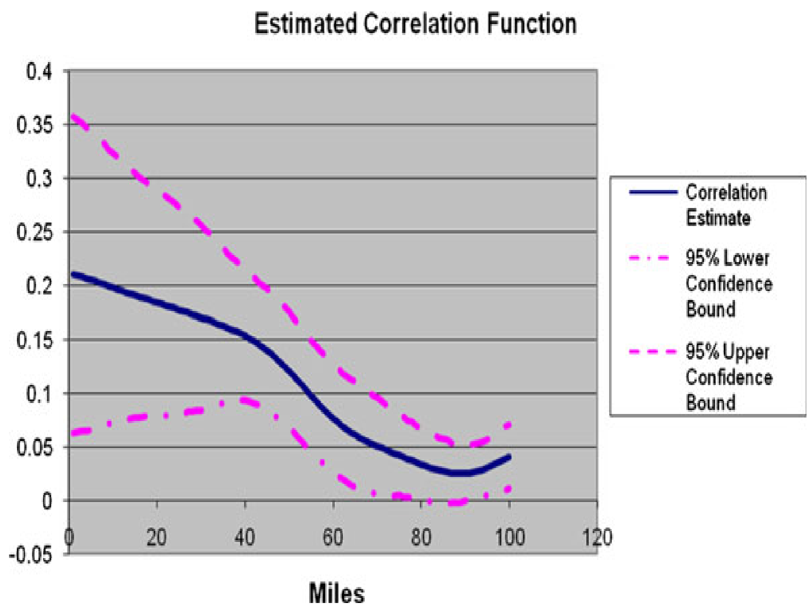

Figure 1: Estimated Correlation Function

mean sample probability for subset j are as presented in equation (4), the relevant decomposition of $\widehat{p}_j - p_j$ is as presented in equation (5), and its first-order Taylor series approximation is as presented in equation (6). It still is the case that the last term disappears even though the random effects are correlated. Then, the asymptotic distribution of $\widehat{p}$ is

$$\sqrt{\widehat{n}}\left(\widehat{p}-p\right) \sim N\left(0,\Omega_p\right)$$

with

$$\begin{aligned} & \Omega_{pjk} \qquad (8) \\ = \; & \widehat{n}E\left[\frac{1}{\widehat{n}_j\sqrt{\widehat{n}}}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)-\frac{\sigma_w w_j}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)+e_j\right]\cdot \\ & \left[\frac{1}{\widehat{n}_k\sqrt{\widehat{n}}}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)-\frac{\sigma_w w_k}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)+e_k\right]' \\ = \; & E\left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)-\frac{\sigma_w w_j}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)+e_j\right]\cdot \\ & \left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i'\sqrt{\widehat{n}}\left(\widehat{\beta}-\beta\right)-\frac{\sigma_w w_k}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)+e_k\right]' \end{aligned}$$

$$
\begin{aligned}
&= E\left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right] \\
&\quad +\sigma_w^2\rho\left(d_{jk}\right)\left(\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)\right)\left(\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)\right)+Ee_je_k'
\end{aligned}
$$

$$
\begin{aligned}
&= \left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right] \\
&\quad +\sigma_w^2\rho\left(d_{jk}\right)\left(\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)\right)\left(\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)\right) \\
&\quad +1\left(j=k\right)\frac{1}{n_j^2}\sum_{i\in J_j}\Phi\left(x_i\beta\right)\left(1-\Phi\left(x_i\beta\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\approx \left[\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)x_i'\right]\Omega_\beta\left[\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)x_i\right] \\
&\quad +\sigma_w^2\rho\left(d_{jk}\right)\left(\frac{1}{\widehat{n}_j}\sum_{i\in\widehat{J}_j}\phi\left(x_i\beta\right)\right)\left(\frac{1}{\widehat{n}_k}\sum_{i\in\widehat{J}_k}\phi\left(x_i\beta\right)\right).
\end{aligned} \tag{9}
$$

5 Wald Test

Our goal is to test H_0^*, i.e., whether subset variation in p_j is meaningful. Define

$$
B=\begin{pmatrix}
1 & -1 & 0 & \cdots & 0 & 0 \\
0 & 1 & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & -1
\end{pmatrix}.
$$

Then, H_0^* can be written as

$$
H_0 : Bp = 0 \quad \text{vs} \quad H_A : Bp \neq 0.
$$

The Wald test statistic is

$$
W=\left(Bp\right)'\Omega_p^{-1}\left(Bp\right)\sim\chi_{N-1}^2 \tag{10}
$$

under H_0. One could use a parametric bootstrap method to approximate the small-sample distribution of the Wald statistic if desired.

6 Empirical Example

The Wald test statistic derived above can be used to estimate statistical significance of parameters or functions of interest in any setting containing area-specific dependent random effects. In this section, we present an example, Johnson et al. (2016a). They estimate the average mental illness prevalence rates across sub-regions of Virginia in the presence of spacially correlated region-specific random effects.

Johnson et al. (2016a) use a national data set with information on the mental illness and demographic characteristics of each sample member along with a local data set, missing data on mental illness but including the same demographic information as the national dataset. The national data set, the 1995 National Health Interview Survey (NHIS), surveys 95091 individuals nationally and includes self-reported mental illness as well as age, race, sex, marital status, education, income, and physical health problems. In addition, it includes sampling information allowing one to determine if two sample observations are from the same (small) geographic area. However, the areas are not labeled in the data; thus they can not be matched to geographic information in the local data set.

The local data set, the 2012 American Community Survey (ACS), surveys approximately 1% of the U.S. population. The survey does not include mental health variables but does include age, race, sex, marital status, education, income, and a set of physical health indicators distinct from the indicators included in the NHIS. It also includes the Public Use Microdata Area (PUMA) where each sample member lives.

The empirical strategy consists of four steps:

1. They estimate the parameters of a random-effects probit model of possessing a mental illness on observable characteristics using the national NHIS data set. In addition to the personal characteristic parameters, they estimate the variance of area-specific random effects, σ_w in equation (7), for the latent variable y_i^* determining the outcome variable y_i.

2. They simulate individual characteristics that are unavailable in the ACS data set but are available in the NHIS data set using observed data in both the ACS and NHIS following Stern (2014). These simulated ACS variables, together with the observed ACS variables, fully constitute the set of personal characteristics used in the random effects probit regression in (1).

3. They construct small regions within Virginia, called composite community service boards (CCSBs), corresponding to the provision of public mental health services and disaggregate the ACS observations by CCSB.

4. They use the probit estimates from (1) to estimate the local mental illness prevalence in each CCSB in Virginia using the observed data for each individual in the ACS and simulated characteristic data from (2) for each

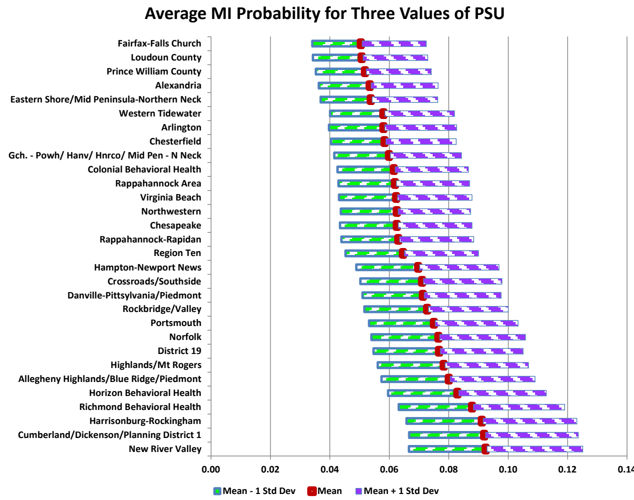

Figure 2: Average MI Probability for Three Values of PSU

> individual in the ACS. Critically, however, they do not know which area-specific random effects, estimated in (1), are associated with each CCSB. Thus, Johnson et al. (2016a) estimate the mental illness prevalence in each CCSB for a range of potential values of the CCSB-specific random effect. Specifically, they estimate the mental illness prevalence in each CCSB for CCSB-specific random effects of 0 and $\pm\sigma_w$.

The range of mental illness prevalences for each CCSB are shown in Figure 2.[6] Under the assumption that there are no CCSB-specific random effects, the local mental illness prevalence varies across CCSBs from 5% to 9%. However, $\sigma_w = 0.201$, which is large. Since Johnson et al. (2016a) do not know the value of any CCSB-specific random effect, they cannot determine whether there is meaningful variation in mental illness propensity across CCSBs due to differences in the joint distribution of observable characteristics that affect mental health. Specifically, note that, in Figure 2, the $-\sigma_w$ propensity for the CCSB with the highest deterministic component of mental illness propensity (New River Valley at 6.7%) is lower than the σ_w propensity for the CCSB with the lowest deterministic component of mental illness propensity (Fairfax-Fall Church at 7.2%). As such, the authors cannot reject a null hypothesis that the distribution of prevalence is constant across CCSBs. If CCSBs with low deterministic components of mental illness prevalences have high CCSB-specific random effects, then mental illness prevalence could be close to constant across CCSBs. Alternatively, if CCSBs with low deterministic components of mental illness prevalences have low CCSB-specific random effects, then the range of mental illness prevalence across counties could be much greater than suggested

[6] This figure is Figure 3 in Johnson et al. (2016a).

by the means in Figure 2.

However, some details about the correlation of CCSB-specific random effects across CCSBs are available. Stern et al. (2010) estimate the correlation of area-specific random effects for mental illness prevalence across small areas based on the distance between them.[7] They find that area-specific random effects are highly correlated for areas that are no more than 40 miles apart. As the distance continues to rise, the correlation in area-specific random effects for mental illness prevalence begins to fall. Once the distance rises to more than 70 miles, area-specific random effects become fairly uncorrelated across areas.

To test the hypothesis that CCSBs have statistically significant differences in mental illness prevalences, a researcher must generate a test statistic that takes into account these CCSB-specific random effects and their correlation across CCSBs. Following Section 5, we construct such a Wald test using the covariance matrix between CCSB-specific random effects inferred from Stern et al. (2010). The test statistic is $\chi^2_{29} = 51.74$ with a p-value of 0.006. Thus, we reject the null hypothesis that there is no variation in mental illness propensity across CCSBs in Virginia.

7 Potential Use in Other Papers

In this section, we consider how our proposed methodology for testing H_0^* could be used in other papers in the literature. The discussion is not exhaustive but still suggestive of the possibilities.

Alderman et al. (2002) and Elbers, Lanjouw, and Lanjouw (2003) estimate a model of income inequality and allow it to vary across small areas. They recognize that there might be important spatial correlation although of a kind more consistent with a model like

$$y_{ijk} = \mu\left(x_{ijk}\right) + \varepsilon_i + \delta_{ij} + \eta_{ijk} \tag{11}$$

where y_{ijk} is the income of person k from village j of region i, x_{ijk} is a vector of observed explanatory variables, and $\left(\varepsilon_i, \delta_{ij}, \eta_{ijk}\right)$ are random effects specific to region, village, and person, respectively. Elbers, Lanjouw, and Lanjouw (2003) fail to reject the null hypothesis that $Var\left(\delta_{ij}\right) = 0$. But this is obviously affected by the degree of spatial correlation as in Johnson et al. (2016).[8] Neither Alderman et al. (2002) or Elbers, Lanjouw, and Lanjouw (2003) test the analog of H_0^*, but both can do so straightforwardly using equation (10) where elements of Ω_p are defined in equation (9) with $\rho\left(\cdot\right) = 0$.

Leroux, Lei, and Breslow (1999) and MacNab (2004), among other papers, use a covariance structure, called a lattice model, allowing only for nonzero covariance for adjacent areas. While this might be a reasonable specification

[7] See Figure 1 above.

[8] Elbers, Lanjouw, and Lanjouw (2003) do not provide enough detail for us to perform the test ourselves.

of the covariance structure, it seems quite restrictive.[9] However, our testing methods still apply. Given sufficient GIS data on location, one can construct an adjacency matrix. Then, the elements of Ω_p in equation (9) can be restricted to zero for non-adjacent areas and can be estimated with a parsimonious specification of the adjacency effect. Construction of the Wald statistic in equation (10) is then straightforward.

Twigg, Moon, and Jones (2000) specify a model of smoking and alcohol-drinking behavior in England but do not allow for a small area-specific effect, much less spatial correlation in the effects. Choy et al. (2008) do the same for the bacteria associated with stomach ulcers. It seems likely that such an effect is important, could be included in the estimation procedure, and would affect results in both papers. Once included and estimated, the Wald test would be easy to compute.

Legler et al. (2002) estimate a model of mammogram usage, similar in structure to equation (11) though using aggregated data. They include a small-area random effect but do not allow for any correlation across areas. The authors do not report any results corresponding to the random effect. Mendez-Luck et al. (2007) estimate a model for asthma. They refer to some results consistent with important random effects but provide no estimate information nor statistical tests of any kind. However, it is likely that asthma prevalence has spatially correlated random effects which are important to include in any meaningful public health analysis and would not be difficult to include.

Banerjee, Wall, and Carlin (2003) construct a survival model for infant death with conditional hazard function $h\left(t \mid x_{ij}\beta, \varepsilon_i\right)$ where x_{ij} is a vector of observed, exogenous covariates for person j in small area i and ε_i is a small-area random effect. They consider both an isotropic specification and a lattice specification for the covariance matrix of ε and estimate the models using Bayesian methods. Later, they show maps displaying median random effects. Our method provides a classical test of the significance of the variation due to x_{ij} across small areas. Before proceeding, one must choose a measure of variation V_i across small areas; the complication is that $h\left(\cdot \mid \cdot\right)$ is a function with no clear way to summarize its effect. Some possibilities are a) $V_i = x_{i\cdot}\widehat{\beta}$ where $x_{i\cdot}$ is the weighted sample average of $\left\{x_{ij}\right\}_j$ and $\widehat{\beta}$ is a consistent estimator of β,[10] b) $V_i = h\left(t^* \mid x_{i\cdot}\beta, 0\right)$ for some t^* of significance, c) $V_i = h\left(t_k^* \mid x_{i\cdot}\beta, 0\right)$ for a finite vector $\left(t_1^*, t_2^*, .., t_K^*\right)$, or d) V_i is a functional of $h\left(\cdot \mid \cdot\right)$ such as

$$\begin{aligned} V_i &= E\left(T \mid x_{i\cdot}, \varepsilon_i = 0\right) = \int t dF_t\left(t \mid x_{i\cdot}\right) \\ &= \int th\left(t \mid x_{i\cdot}\beta, 0\right) S\left(t \mid x_{i\cdot}\beta, 0\right) dt \end{aligned}$$

[9] On the other hand, Banerjee, Wall, and Carlin (2003) find similar results for this specification and one closer to the specification we use.

[10] Note that, in this example, as is true of the example in Section 6, because the outcome of interest is a nonlinear function of the inputs, proper specification and estimation of the covariance matrix of the random effects must be performed simultaneously with the rest of the model (Heckman and Singer, 1984).

where $F_t(\cdot \mid x_{i\cdot})$ is the distribution of (random) time of death T and

$$S(t \mid x_{i\cdot}\beta, 0) = \Pr[T > t \mid x_{i\cdot}\beta, \varepsilon_i = 0] = \exp\left\{-\int_0^t h(s \mid x_{i\cdot}\beta, 0)\, ds\right\},$$

or

$$V_i = \widetilde{E}(T) = \int E(T \mid x, \varepsilon_i = 0)\, dF_{x_i}(x)$$

where $F_{x_i}(x)$ is the joint distribution of x_i in small area i. For each choice, in order to use the Wald statistic in equation (10), one must first compute $\partial V_i / \partial x$ and $\partial V_i / \partial \varepsilon_i$ to use in the construction of the elements of the analog to Ω_p in equation (9).[11]

Malec and Müller (2008) consider a model with a completely flexible (nonparametric) specification of the spatial covariance matrix of random effects. Their model requires data on location, and asymptotic results require the number of observations in each location to go to infinity. They use Bayesian methods and do not discuss asymptotics. Our methodology is not a good fit for their problem.

References

[1] Alderman, H. M. Babita, G. Demombynes, N. Makhatha, and B. Ozler (2002). "How Low Can You Go? Combining Census and Survey Data for Mapping Poverty in South Africa." *Journal of African Economics.* 11:169-200.

[2] Banerjee, Sudipto, Melanie Wall, and Bradley Carlin (2003). "Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota." *Biostatistics.* 4(1): 123–142.

[3] Biggeri, A., F. Divino, A. Frigessi, A. Lawson, D. Böhning, E. Lesaffre, and J. Viel (1999). "Introduction to Spatial Models in Ecological Analysis." *Disease Mapping and Risk Assessment for Public Health.* (eds.) A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J. Viel, and R. Bertollini, Wiley, New York: 181–192.

[4] Choy, M., P. Switzer, C. De Martel, and J. Parsonnet (2008). "Estimating Disease Prevalence Using Census Data." *Epidemiology & Infection.* 136: 1253–1260.

[5] Citro, C., and G. Kalton (2000). *Small Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Washington, DC: National Academy Press.

[6] Congdon, Peter (2006). "Estimating Population Prevalence of Psychiatric Conditions by Small Area with Applications to Analysing Outcome and Referral Variations." *Health & Place.* 12: 465–478.

[11] This was done for the model discussed in this paper in equation (8).

[7] Congdon, Peter (2009). "A Multilevel Model for Cardiovascular Disease Prevalence in the US and Its Application to Micro Area Prevalence Estimates." *International Journal of Health Geographics*. 8(6).

[8] Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley and Sons, New York.

[9] Elbers, Chris, T. Fujii, P. Lanjouw, B. Ozler, and W. Yin (2007). "Poverty Alleviation through Geographic Targeting: How Much Does Disaggregation Help?" *Journal of Development Economics*. 83: 198-213.

[10] Elbers, Chris, Jean Lanjouw, and Peter Lanjouw (2003). "Micro-Level Estimation of Poverty and Inequality." *Econometrica*. 71(1): 355-364.

[11] Ellis, Alan, Thomas Konrad, Kathleen Thomas, and Joseph Morrissey (2009). "County-Level Estimates of Mental Health Professional Supply in the United States." *Psychiatric Services*. 60(10): 1315–1322.

[12] Ferrándiz, J., A. López, and P. Sanmartin, (1999). "Spatial Regression Models in Epidemiological Studies." *Disease Mapping and Risk Assessment for Public Health*. (eds.) A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J. Viel, and R. Bertollini, Wiley, New York: 203–215.

[13] Ghosh, M., K. Natarajan, T. Stroud, and B. Carlin (1998). "Generalized Linear Models for Small-Area Estimation." *Journal of the American Statistical Association*. 93: 273–282.

[14] Ghosh, M. and J. Rao (1994). "Small Area Estimation: An Appraisal." *Statistical Science*. 9: 55-93.

[15] Heckman, J. and B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*. 52(2): 271-320.

[16] Hudson, Christopher (2009). "Validation of a Model for Estimating State and Local Prevalence of Serious Mental Illness." *International Journal of Methods in Psychiatric Research*. 18(4): 251–264.

[17] Jarjoura, David, Gary McCord, Charles Holzer, and Timothy Champney (1993). "Implementing Indirect Needs-Assessment Synthetic Estimation of the Distribution of Mentally Disabled Adults for Allocations to Ohio's Mental Health Board Areas." *Evaluation and Program Planning*. 16(4): 305-313.

[18] Johnson, William, Michael LaForest, Brett Lissenden, and Steven Stern (2016a). "Variation in Mental Health Illness and Provision of Public Services." Unpublished manuscript.

[19] Kessler, Ronald, Patricia Berglund, Ellen Walters, Philip Leaf, Anthony Louzis, Martha Bruce, Robert Friedman, René Grosser, Cille Kennedy, Timothy Kuehnel, Eugene Laska, Ronald Manderscheid, William Narrow, Robert Rosenheck, and Max Schneier (1998). "A Methodology for Estimating the 12-Month Prevalence of Serious Mental Illness." *Mental Health, United States, 1998*. US Department of Health and Human Services, SAMHSA.

[20] Konrad, Thomas, Alan Ellis, Kathleen Thomas, Charles Holzer, and Joseph Morrissey (2009). "County-Level Estimates of Need for Mental Health Professionals in the United States." *Psychiatric Services*. 60(10): 1307-1314.

[21] Legendre, P., and L. Legendre (1998). *Numerical Ecology.* Elsevier Science, Amsterdam.

[22] Legler, Julie, Nancy Breen, Helen Meissner, Don Malec, and Cathy Coyne (2002). "Predicting Patterns of Mammography Use: A Geographic Perspective on National Needs for Intervention Research." *Health Services Research.* 37(4): 929–947.

[23] Leroux B., X. Lei, and N. Breslow (1999). "Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence." in (eds.) M. Halloran and D. Berry. *Statistical Models in Epidemiology, the Environment and Clinical Trials*. Springer-Verlag: New York:135-178.

[24] Lichstein, Jeremy, Theodore Simons, Susan Shriner, and Kathleen Franzreb (2002). "Spatial Autocorrelation and Autoregressive Models in Ecology." *Ecological Monographs*. 72(3): 445–463.

[25] MacNab, Ying (2004). "Bayesian Spatial and Ecological Models for Small-Area Accident and Injury Analysis." *Accident Analysis and Prevention*. 36: 1019–1028.

[26] Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

[27] Malec, D., J. Sedransk, C. Moriarity, and F. LeClere (1997). "Small Area Inference for Binary Variables in the National Health Interview Survey." *Journal of the American Statististical Association*. 92: 815–826.

[28] Malec, Donald and Peter Müller (2008). "A Bayesian Semi-Parametric Model for Small Area Estimation." *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*. 3: 223–236.

[29] Mendez-Luck, Carolyn, Hongjian Yu, Ying-Ying Meng, Mona Jhawar, and Steven Wallace (2007). "Estimating Health Conditions for Small Areas: Asthma Symptom Prevalence for State Legislative Districts." *Health Services Research.* 42(6): 2389–2409.

[30] Messer, S., X. Liu, C. Hoge, D. Cowan, and C. Engel (2004). "Projecting Mental Disorder Prevalence from National Surveys to Populations-of-Interest: An Illustration Using ECA Data and the US Army." *Social Psychiatry and Psychiatric Epidemiology.* 39: 419–426.

[31] Opsomer, J., G. Claeskens, M. Ranalli, G. Kauermann, and F. Breidt (2008). "Non-parametric Small Area Estimation Using Penalized Spline Regression." *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 70(1): 265-286.

[32] Prasad, N., and J. Rao (1999). "On Robust Small Area Estimation Using a Simple Random Effects Model." *Survey Methodology.* 25: 67–72.

[33] Rao, J. (1999). "Some Recent Advances in Model-Based Small Area Estimation." *Survey Methodology.* 25: 175-186.

[34] Rossi, R., D. Mulla, A. Journel, and E. Franz (1992). "Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence." *Ecological Monographs.* 62(2): 277-314.

[35] Stern, Steven (2014). "Estimating Local Prevalence of Mental Health Problems." *Health Services and Outcomes Research Methodology.* 14: 109-155.

[36] Stern, Steven, Elizabeth Merwin, Emily Hauenstein, Ivora Hinton, Virginia Rovnyak, Melvin Wilson, Ishan Williams, and Irma Mahone (2010). "The Effects of Rurality on Mental and Physical Health." *Health Services and Outcomes Research Methodology.* 10(1): 33-66.

[37] Tarozzi, Alessandro, and Angus Deaton (2009). "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas." *Review of Economics and Statistics.* 91(4): 773–792.

[38] Twigg, L., G. Moon, and K. Jones (2000). "Predicting Small-Area Health-Related Behaviour: A Comparison of Smoking and Drinking Indicators." *Social Science & Medicine.* 50: 1109–1120.