Supplement to "Macroeconomic effects on mortality revealed by panel analysis with nonlinear trends"

Edward L. Ionides[1], **Zhen Wang**[1] **and José A. Tapia Granados**[2]

December 28, 2012

[1] Department of Statistics and [2]Institute for Social Research,
The University of Michigan, Ann Arbor, Michigan, USA.
email: ionides@umich.edu, wzhen@umich.edu, jatapia@isr.umich.edu

Accepted for publication in *Annals of Applied Statistics*

Supplementary Content

S1 Interpretation of detrending choices

The frequency response function, $A(\omega)$, for a filter gives the fraction of the frequency component of the time series which is retained by the filter at frequency ω (Shumway and Stoffer, 2006). The frequency response function of the Hodrick-Prescott filter is

$$A_{\mathrm{HP}}(\omega, \lambda) = \frac{4\lambda\{1 - \cos(\omega)\}^2}{1 + 4\lambda\{1 - \cos(\omega)\}^2},$$

where λ is the smoothing parameter (King and Rebelo, 1993). The frequency response function of the difference filter is

$$A_{\mathrm{D}}(\omega) = 1 - \exp\{-i\omega\},$$

which is complex valued since the filter is not symmetric (Shumway and Stoffer, 2006). Figure S-1 graphs the frequency response function for the Hodrick-Prescott filter (at $\lambda = 100$ and $\lambda = 6.25$) and the difference filter. This figure shows how the choice of $\lambda = 6.25$ effectively removes cyclical components with period of more than around 8 years.

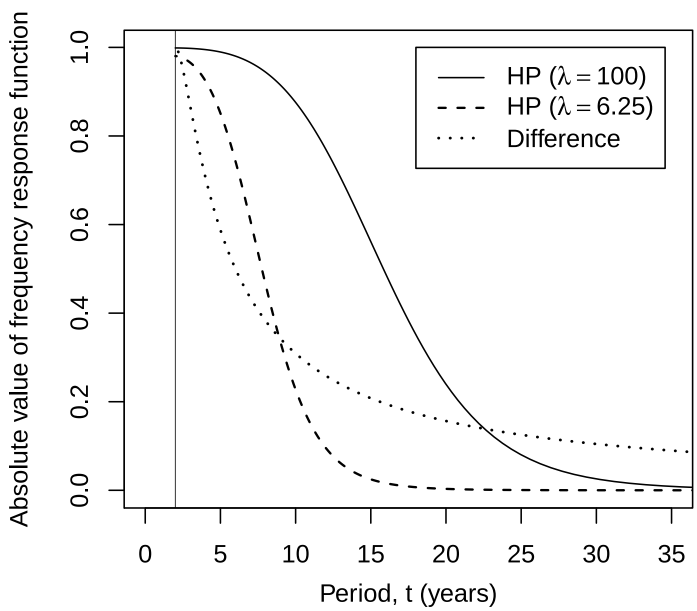

Figure S-1: Frequency response functions, plotted as a function of the period of the oscillation ($t = 2\pi/\omega$). The solid and dashed lines show $A_{\mathrm{HP}}(2\pi/t, \lambda)$ for $\lambda = 100$ and $\lambda = 6.25$. The vertical line at $t = 2$ corresponds to the Nyquist frequency: annual data cannot resolve oscillations with shorter period than 2 years. The dotted line shows $|A_{\mathrm{D}}(2\pi/t)|/2$. Since $A_{\mathrm{D}}(\pi) = 2$, the difference filter actually inflates frequency components close to the Nyquist frequency and the rescaling factor of 2 is appropriate for comparison.

S2 Data analysis for additional detrending choices

The consequences for our data analysis of the choice of the Hodrick-Prescott smoothing parameter, λ, are shown in Table S-1. The results for $\lambda = 6.25$ are generally intermediate between $\lambda = 100$ and analyses based on differencing. This might be expected from Figure S-1, since the graph of the frequency response function for the $\lambda = 6.25$ lies between $\lambda = 100$ and the difference filter for oscillatory periods between 2.3 and 8.8 years.

Ordinary least squares (the usual, practical and convenient methodology) is statistically efficient and appropriate when the residuals have weak or negligible sample correlation. Figure S-2 shows that small values of the Hodrick-Prescott smoothing parameter, such as $\lambda = 6.25$, result in considerable negative temporal autocorrelation of the residuals. This can be expected to make the corresponding OLS estimates inefficient. By contrast, $\lambda = 100$ serves approximately as a prewhitening filter.

Table S-1: Percentage increase in mortality associated with a unit increase in the state unemployment rate for model $HP1_{\lambda}$ at four choices of λ, and for model D1.

Model	$HP1_{400}$	$HP1_{100}$	$HP1_{25}$	$HP1_{6.25}$	D1
Total	-0.35 *** *** ***	-0.33 *** *** ***	-0.30 *** *** **	-0.24 *** *** *	-0.16 * *
20-44	-0.47 ** *	-0.47 ** *	-0.58 *** ** *	-0.73 *** ** *	-0.54 * * +
45-64	-0.26 *** * *	-0.22 ** * *	-0.15	-0.14	-0.13
65+	-0.24 *** *** **	-0.25 *** *** *	-0.24 *** *** *	-0.16 * +	-0.03
Cardiovascular disease	-0.25 *** ** *	-0.24 *** ** *	-0.21 * * +	-0.14	-0.06
Ischemic heart disease	-0.55 *** ** *	-0.58 *** ** *	-0.47 *** ** *	-0.28 + +	-0.14
Cancer	0.06	0.04	0.03	0.05	0.13
Respiratory disease	-0.54 *** *** *	-0.71 *** *** **	-0.82 *** *** **	-0.69 ** ** *	-0.37
Other infectious disease	-0.88 *	-0.89 * +	-1.26 ** ** *	-1.72 *** ** **	-1.14 * * *
Traffic injury	-2.56 *** *** ***	-2.11 *** *** ***	-1.74 *** *** ***	-1.44 *** *** ***	-1.48 *** *** ***
Suicide	0.91 *** * *	0.77 ** + +	0.76 * + +	0.80 * + +	0.94 * + *
Homicide	-0.65	-0.41	-0.34	-0.74	-1.02

Columns represent models, as described in equation (1) and Table 1 of the main text. Rows represent mortality categories. Table entries are estimates of 100α, using OLS with states weighted by the square root of the state population. Statistical significance is shown using standard OLS errors (black symbols, top row), error estimates clustered by state (gray symbols, middle row) and error estimates of Cameron et al. (2011, Section 2.2) with two-way clustering by state and year (gray symbols, bottom row; red in electronic version). *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, $^{+}$ $P < 0.1$.

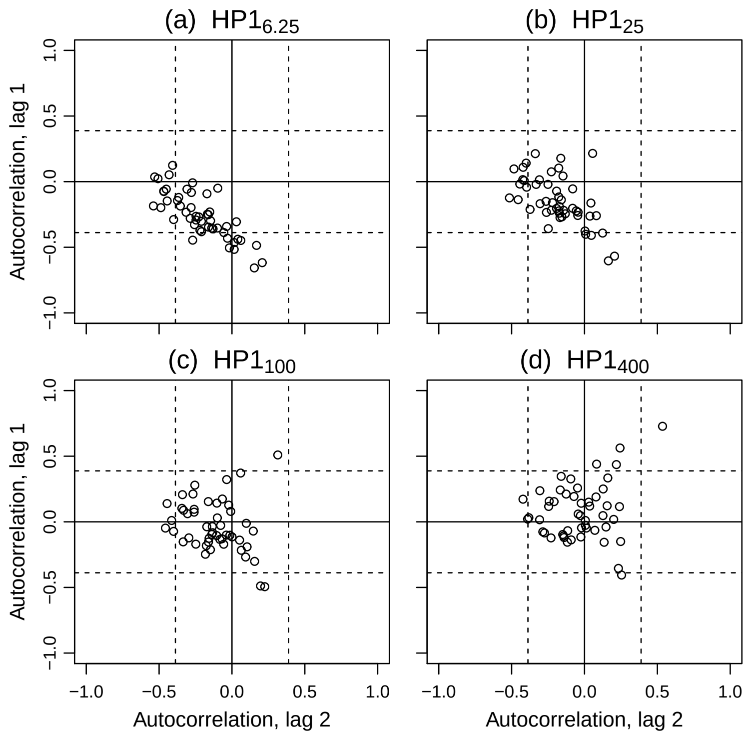

Figure S-2: Autocorrelation of the residuals from fitting model HP1$_\lambda$ for four values of the Hodrick-Prescott smoothing parameter λ. Points show the sample autocorrelation for each state, at lag 1 and lag 2. The horizontal dashed lines are at $\pm t_{n-2}\{n-2+t_{n-2}^2\}^{-1/2}$ where t_{n-2} is the 97.5 percentile of the t distribution on $n-2$ degrees of freedom, and $n = 26$. If the residual series were temporally uncorrelated, approximately 95% of the points should lie between the dashed lines (Moore and McCabe, 1999, Section 10.2). The vertical dashed lines are constructed similarly, with $n = 25$.

S3 Prewhitening for selecting the Hodrick-Prescott smoothing parameter

A classic time series regression approach to efficient statistical inference in the presence of autocorrelated errors is to apply a linear prewhitening filter, i.e., to apply a linear transformation to all the variables in the regression so that the residuals are approximately uncorrelated (Shumway and Stoffer, 2006). Empirically, for our data analysis, a value of $\lambda = 6.25$ leads to considerable negative autocorrelation in the residuals (Figure S-2) whereas $\lambda = 100$ is approximately a prewhitening filter.

Suppose that one wished to estimate the regression coefficients in equation (1) of the main text, using Hodrick-Prescott detrending with $\lambda = 6.25$. Let $f_{6.25}(\omega)$ be the power spectrum of the resulting residuals. Then, the power spectrum of the residuals when detrending with $\lambda = 100$ is approximately

$$f_{100}(\omega) \approx f_{6.25}(\omega) \frac{|A_{\mathrm{HP}}(\omega, 100)|^2}{|A_{\mathrm{HP}}(\omega, 6.25)|^2}, \tag{S1}$$

where $A_{\mathrm{HP}}(\omega, \lambda)$ is the frequency response function of the HP filter as in Section S1. Equation (S1) would be exact if the choice of prewhitening filter had no consequences for parameter estimation; the approximation results from the effect of the prewhitening filter on the estimated parameter values. So far as the approximation in (S1) is valid, if $\lambda = 100$ is an effective prewhitening filter for the original data then $A_{\mathrm{HP}}(\omega, 100)\left[A_{\mathrm{HP}}(\omega, 6.25)\right]^{-1}$ gives the frequency response function for an effective prewhitening filter for the data detrended using $\lambda = 6.25$. This prewhitening filter can be implemented by undoing the $\lambda = 6.25$ detrending and applying $\lambda = 100$. In other words, if one wants to estimate the regression coefficients in equation (1) based on detrending with $\lambda = 6.25$, it is reasonable to use instead the estimates resulting from using $\lambda = 100$.

The conclusion of the previous paragraph is perhaps counter-intuitive. From a theoretical perspective, the paradox can be resolved by noting that any linear detrending method gives unbiased parameter estimates in the context of the usual linear model that we present in equation (1) of the main text, so long as it successfully removes deterministic trends. Prewhitening then becomes the key criterion for efficient parameter estimation. The observation that $\lambda = 100$ is an empirically effective prewhitening filter suggests that it does a respectable job of removing deterministic trends: if present, these would usually be associated with high power at low frequencies.

From a practical perspective, this theoretical resolution is not entirely satisfactory. Likely, a model such as equation (1) is not equally valid across all spatiotemporal scales and so detrending methods which emphasize different aspects of the data may not in fact be estimating the same quantity. Unless there are reasons for restricting attention to specific temporal scales, this practical consideration justifies our decision to present results for a range of detrending methodologies.

The theoretically privileged position of the prewhitening choice $\lambda = 100$ does seem to have some practical consequences: it also happens to result in the clearest statistical evidence for procyclical mortality in our data analysis (e.g., the greatest number of significance stars in Table 3 of the main text). However, this does not rule out the possibility that some effects may be more clearly apparent with an alternative analysis that may also have properly justified levels of statistical significance.

Supplementary References

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

King, R. G. and Rebelo, S. T. (1993). Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control*, 17(1–2):207 – 231.

Moore, D. S. and McCabe, G. P. (1999). *Introduction to the Practice of Statistics*. W. H. Freeman and Company, New York, 3rd edition.

Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and Its Applications*. Springer, New York, 2nd edition.