DISCERNING TRENDS IN COMMODITY PRICES

Dimitri Dimitropoulos and Adonis Yatchew

UNIVERSITY OF TORONTO

Workshop on Commodity Super Cycles – Bank of Canada (Ottawa)
April 27-28, 2015

Motivation

- Commodity prices are driven by a complexity of forces:
  - depletion and discovery
  - innovation and obsolescence
  - competition and strategic behaviour
  - geopolitics and conflict
  - growth, development and business cycles.

- Our departure point -- price trends are fundamentally nonparametric.

- This paper applies nonparametric methods to price data on 11 commodities -- 3 hydrocarbons and 8 metals -- for the period 1900-2014.

Nonparametric Modeling of Trend

We draw on three ideas in the nonparametrics literature to construct and estimate parsimonious models:

1. The Partial Linear Model

2. Cross-Validation

3. Shape Similarity

1. The Partial Linear Model

- Reference specification --

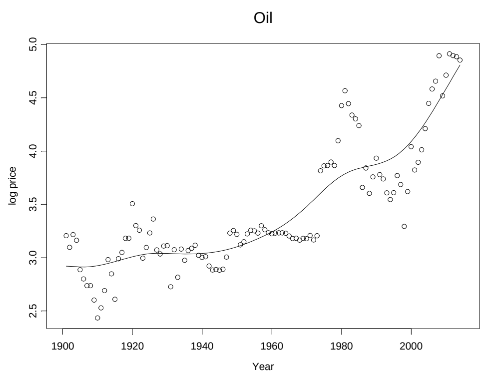
$$y_t = f(t) + z_t\beta + \varepsilon_t$$

  - y_t is the (log) *real* commodity price
  - f is an unknown smooth function of time t
  - z_t is a vector of observable variables, β is a vector of unknown parameters
    - z_t may include macro variables, shift dummies …
  - ε_t is an error term which may be heteroskedastic and/or serially correlated.

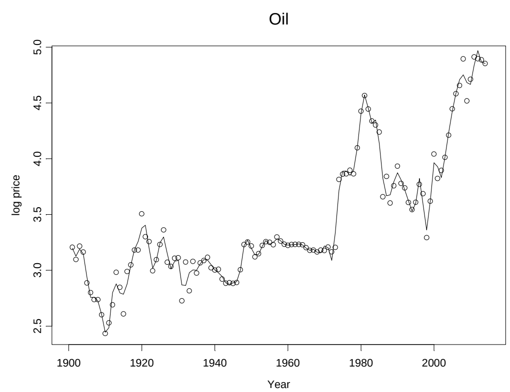
2. Cross-Validation

- Consider a pure nonparametric trend model $y_t = f(t) + \varepsilon_t$. Nonparametric estimators 'smooth' the data.
- How does one select an appropriate degree of smoothing?

Over-smoothing



Under-smoothing



- Minimizing the mean square error $\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{f}(t,\lambda)\right)^2$ by choosing the smoothing parameter λ will not work – one obtains a perfect fit.
- Solution? Cross-validation.

2. Cross-Validation in *iid* Setting

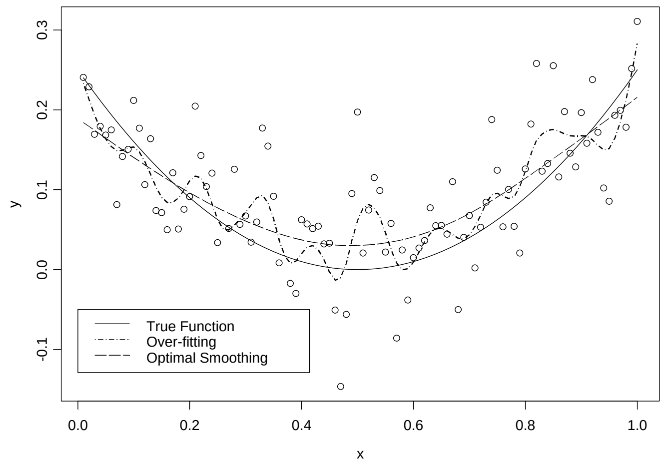
- In the *iid* setting, cross-validation permits optimal selection of the smoothing parameter λ.
  - Fix λ and for each t, estimate $f(t)$ while *omitting* the t-th observation.
  - Repeat the process over a grid of values of λ.
  - Select the value of λ that minimizes mean squared error:

$$\min_{\lambda} \; CV(\lambda) = \tfrac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{f}_{-t}(t,\lambda)\right)^2 .$$

- Essentially, determination of the degree of smoothing is based on the ability to predict 'out of sample'.

2. Cross-Validation in *AR* Setting

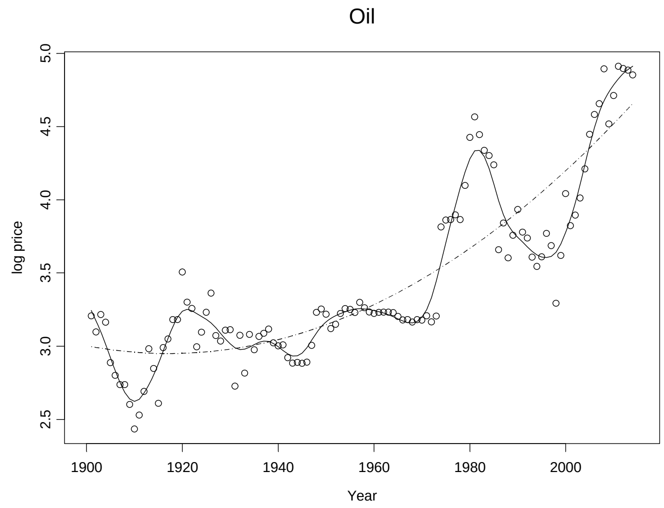
- Nearby observations are not statistically independent.
- Omitting *individual* observations will lead to over-fitting.

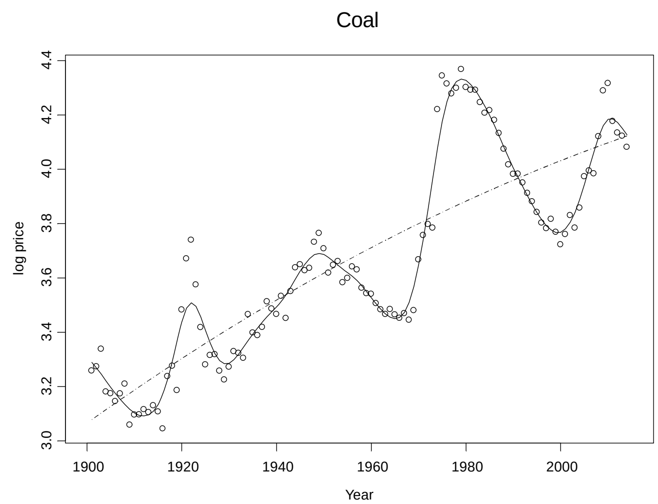


- Solution? Omit observations in a neighborhood *N(t)* of each point:

$$\min_{\lambda} \quad CV(\lambda) = \tfrac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{f}_{-N(t)}(t,\lambda)\right)^2 \quad .$$

3. Shape Similarity





Model Features

1. Our parsimonious specification $y_t = f(t) + z_t\beta + \varepsilon_t$ balances flexibility and precision:
   i. Nonparametric component allows flexible identification of trend.
   ii. But, a pure nonparametric model where *z* variables are also treated nonparametrically is subject to 'curse of dimensionality'.
   iii. Macroeconomic effects are arguably amenable to parametric modeling, e.g., might expect unemployment to have monotone effect on copper prices. Shift variables, which capture 'regime change' may be included in the *z* vector.

2. Cross-validation allows data-driven determination of smoothness parameter.

3. Tools available for assessing/testing shape similarity of trends.

Model Estimation

- $\sqrt{n}$ -consistent estimates of the coefficients on the *z's* can be obtained by regressing the 'detrended' price on the 'detrended' parametric variables

$$y_t - E[y_t \mid t] = \left(z_t - E[z_t \mid t]\right)\beta + \varepsilon_t \ .$$

- Since the residuals are likely heteroskedastic and serially correlated, we report Newey-West standard errors, adapted to the current setting.

- After removing the estimated parametric effects , the trend effect can be estimated by performing nonparametric regression on the model

$$y_t - z_t\hat{\beta} \cong f\left(t\right) + \varepsilon_t \,.$$

- Asymptotic confidence bands can be constructed around the estimate of nonparametric function to gauge precision of estimated trend.

Endogeneity Issues

- In estimating the parametric effects, we have a standard regression model of the form

$$y_t^* = z_t^* \beta + \varepsilon_t$$

where $y_t^* = y_t - E[y_t \mid t]$ and $z_t^* = z_t - E[z_t \mid t]$

- However, in modeling oil prices, or more generally hydrocarbon prices, certain macroeconomic variables (e.g., unemployment) are likely to be correlated with the residual.

- In these cases we apply instrumental variable estimation to the partial linear model.

Data

- Prices of 11 commodities: 3 hydrocarbons (oil, natural gas, coal) and 8 metals (copper, nickel, zinc, iron, tin, silver, lead, aluminium) for the period 1901 to 2014.

- Main source is Manthy (1978), from which we obtain the commodity price series and the wholesale price index prior to 1973.

- Additional sources are used to update the data to 2014.
  - The price series for oil, natural gas and coal are augmented using prices published by the US Energy Information Administration (EIA).
  - For metals prices, the US Geological Survey *Historical Statistics for Mineral and Material Commodities in the United States* is used to complete the time series to 2012, and LME prices through to 2014.
  - The wholesale price index is updated using the *Producer Price Index (Commodities)* as published by the U.S. Bureau of Labor Statistics.

Data

- Macroeconomic variables are included to control for short term business cycle dynamics and longer term economic growth trends;
  - growth rate of the Gross Domestic Product (GDP) of 20 OECD countries as well as 6 developing countries.
    - Source: Maddison Project's *Statistics on World Population and GDP 1-2010*, updated to 2014 using World Bank's *World Development Indicators* database.
  - US unemployment rate series constructed from Lesbergott ('57) for the period 1900-42, and from the Bureau of Labor Statistics through to 2014.
  - UK unemployment rate series constructed from Denman and McDonald ('96) for the period 1900-70, and from the Office of National Statistics through to 2014.

- Indicator variables for events that are expected to have influenced commodity prices, e.g., OPEC actions in 1973, U.S. Federal mining and environmental legislation in the late 1960s.

Hotelling and His Descendants

- Hotelling Rule tells us that the real price of an exhaustible resource should be trending upward in the long run.

- However, such predictions are typically rejected by data, as prices are often observed to fall, at least over certain periods of time.

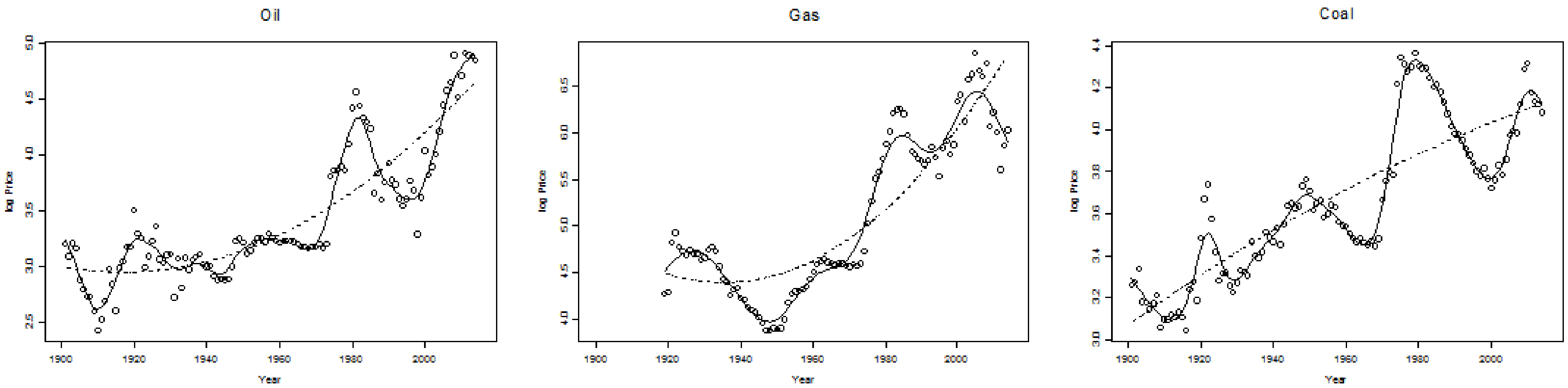
- In attempts to rationalize this phenomenon, various authors have proposed extensions of the basic framework to allow for U-shaped trends, oscillatory trends, structural breaks….

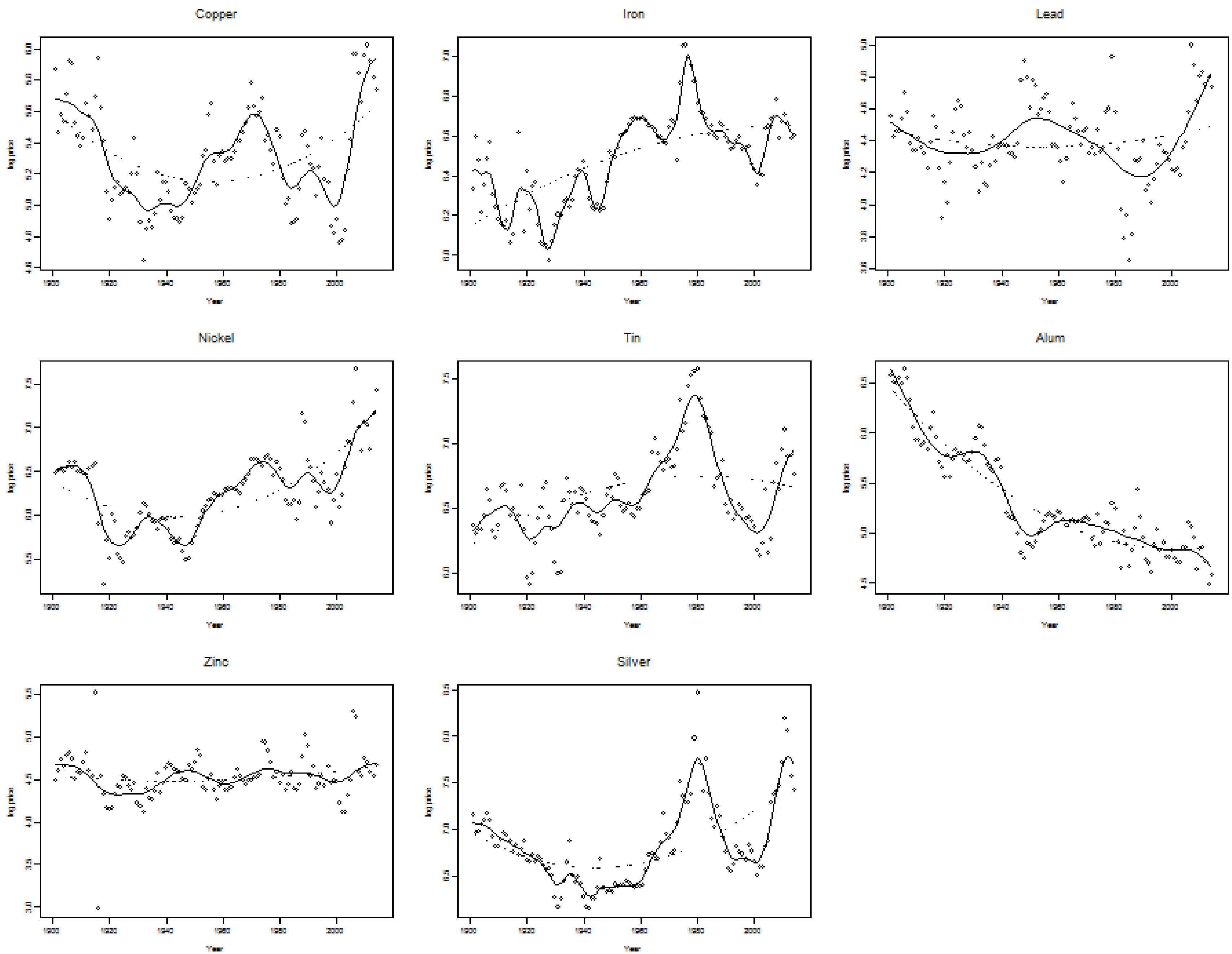
Pure Trend Models

Our analysis of the data begins with testing of a pure quadratic trend model against its nonparametric alternative:

$$H_0: \ y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

$$H_1: \ y_t = f(t) + \varepsilon_t$$

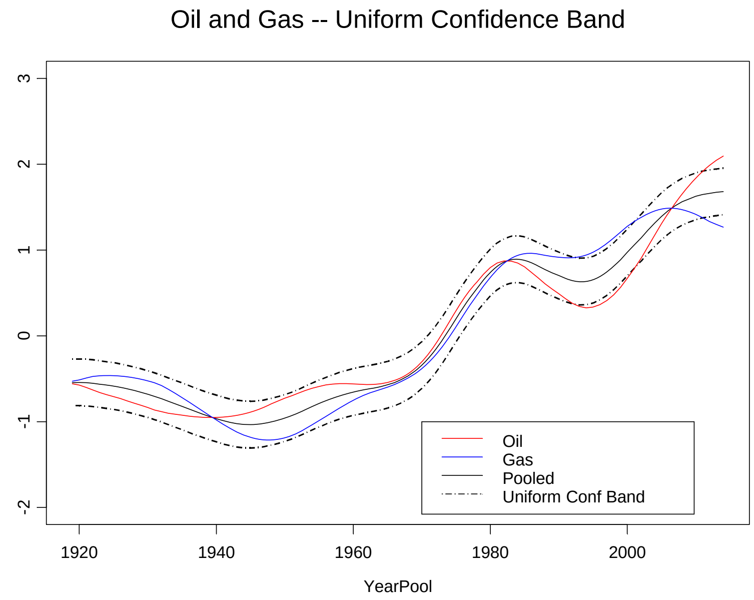



Copper
Iron
Lead
Nickel
Tin
Alum
Zinc
Silver
Year
log price

Pure Trend Models

- In most cases the quadratic trend model is rejected in favour of the nonparametric version, the exceptions being zinc, lead and aluminum.
- Casual inspection suggests degree of shape similarity among various groups: (i) hydrocarbons, (ii) copper-nickel, (iii) iron-tin.
- Aluminum is only commodity that displays a strong secular downward trend.
- We consider more formally whether certain commodities bear shape similarity in their longer term trends by comparing individual trends against a pooled estimate, together with its 95% uniform confidence band.

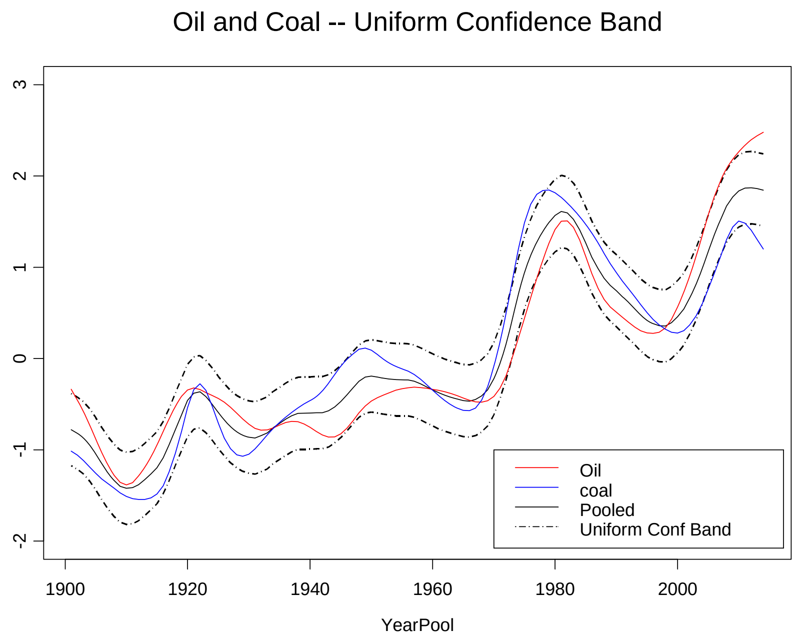
Pure Trend Models -- Shape Similarity



- Both the oil and gas trend estimates lie within the pooled uniform band for most of the data period.

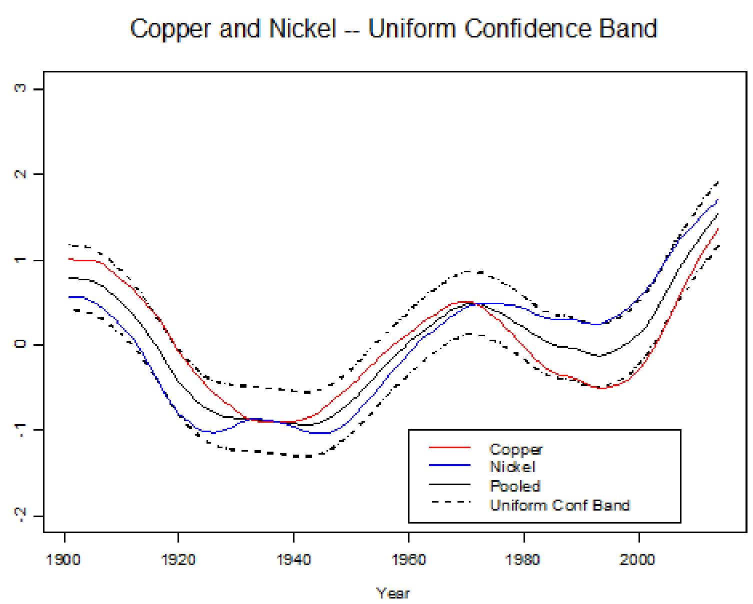
- Beginning around 2009, the individual price trends diverge -- U.S. natural gas prices plummet first, as a result of the fracking revolution, while oil prices remain high until late 2014, at which time they also plunge.

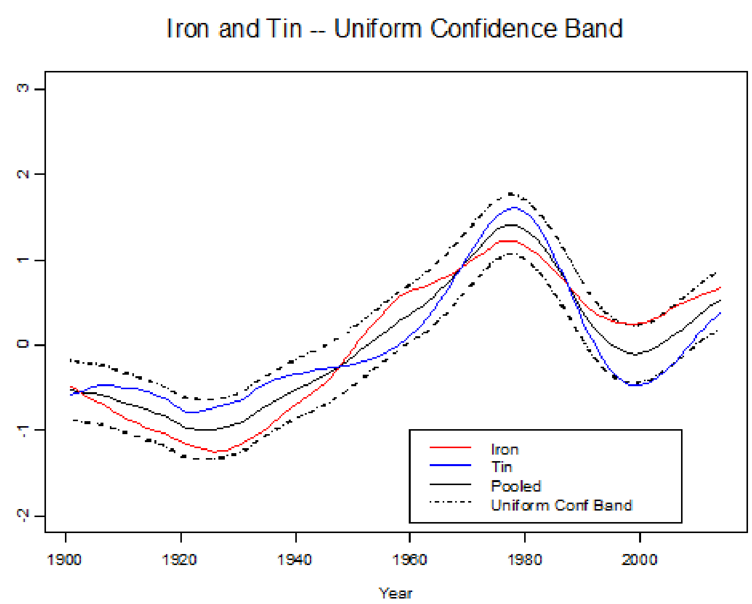
Pure Trend Models -- Shape Similarity



- Both oil and coal trends follow generally similar paths, although there are some differences in timing.
- For most of the 1901-2014 period, oil and coal trends generally lie within the 95% uniform confidence bounds of the pooled trend estimate.

Pure Trend Models -- Shape Similarity





- Metals prices exhibit considerable volatility, but the estimated trend curves for copper and nickel, and iron and tin, show a degree of similarity.
- In both cases, the uniform bounds of the pooled estimate capture the individual trend paths for each of the metals pairs.

Inclusion of Macro Variables

- We now incorporate parametric effects into our models, and estimate the partial linear model of the form:

$$y_t = f(t) + z_t \beta + \varepsilon_t$$

- Smoothness parameters are determined through cross-validation.

Inclusion of Macro Variables – Hydrocarbons

Table 3A: Macro Variable Estimates -- Hydrocarbons						
	Oil (1901-2014)		Natural Gas (1992-2014)		Coal (1946-2014)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
OECD Gr.	-0.0006	0.0093	---	---	---	---
EMERGE Gr.	-0.0006	0.0049	---	---	---	---
US Gr.	---	---	0.0519	0.0163**	-0.0005	0.0013
US Unemp.	---	---	-0.0312	0.0547	-0.0082	0.0047*
US/UK Unemp.	-0.0305	0.0165*	---	---	---	---
OPEC Dummy	0.5354	0.0714**	---	---	0.2758	0.0416**
OPEC Share	---	---	0.1336	0.0801*	---	---
Reg. Dummy	---	---	---	---	0.1457	0.0255**
R^2	0.96 (0.95)		0.94 (0.91)		0.98 (0.96)	
	χ^2_4	p-val	χ^2_4	p-val	χ^2_4	p-val
Joint Sign.	141.0	<0.01	42.7	<0.01	76.6	<0.01
	$N(0,1)$	p-val	$N(0,1)$	p-val	$N(0,1)$	p-val
Quad. Null	11.9	<0.01	7.6	<0.01	8.9	<0.01

Inclusion of Macro Variables – Hydrocarbons

- These exhibit strong joint significance in the models for all three fuels.
- Their inclusion would appear to have a modest impact on total explanatory power. The reason is that in the pure nonparametric model, the trend variable is picking up macro effects. In the partial linear model, the role of trend declines significantly.
- For oil prices,
  - Neither OECD nor emerging economies growth rates have a significant impact.
  - The average US/UK unemployment rate is marginally significant.
  - The change in strategic behaviour by OPEC beginning in late 1973 is estimated to increase oil prices, on average, in excess of 50%.

Inclusion of Macro Variables – Hydrocarbons

- For natural gas prices,
  - As these markets are primarily continental. We use U.S. prices and therefore U.S. growth and unemployment rates.
  - We focus on post 1992 data, as this market went through a period of price constraints for inter-state trade until the early 1990's. (Estimates for 1919-2014 are available.)
  - U.S. growth is a strongly significant driver of continental natural gas prices, while U.S. unemployment is not.
  - Note that an OPEC dummy would not be identified, so in this case we include the OPEC market share -- which is significant at the 10% level.

- For coal prices,
  - We use U.S. growth and unemployment rates.
  - We focus on the post-war period, 1946-2014.
  - U.S. economic growth is not a significant factor in explaining the price series, but unemployment is marginally significant with a negative impact on prices.
  - The OPEC effect is large and strongly significant, increasing coal prices in excess of 25% on average relative to trend.
  - The "Regulatory Dummy" -- which captures the impacts of coal mine safety legislation passed in 1969, and the amendments to the Clean Air Act in 1970 -- is strongly statistically significant, increasing coal prices by an estimated 15%.

Inclusion of Macro Variables – Metals

Table 3B: Macro Variable Estimates -- Metals 1901-2014						
	Copper		Tin		Lead	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
OECD Gr.	0.0085	0.0033**	0.0133	0.0052**	0.0101	0.0043**
EMERGE Gr.	-0.0035	0.0070	0.0034	0.0065	-0.0029	0.0079
US/UK Unemp.	-0.0358	0.0104**	-0.0224	0.0122*	-0.0206	0.0084**
R^2	0.81 (0.77)		0.83 (0.80)		0.49 (0.42)	
	χ^2_4	p-val	χ^2_4	p-val	χ^2_4	p-val
Joint Sign.	20.6	<0.01	16.0	<0.01	14.0	<0.01
	$N(0,1)$	p-val	$N(0,1)$	p-val	$N(0,1)$	p-val
Quad. Null	3.5	<0.01	7.8	<0.01	-1.0	0.84

Inclusion of Macro Variables – Metals

- There is substantial variation in the data-driven smoothing parameter, leading to considerable variation in the goodness-of-fit across models.
- The parametric variables – OECD and emerging economy growth rates, and the US/UK unemployment rate – are jointly significant in 5 out of 8 metals.
- OECD growth rates are significant in the copper and lead equations, while growth in emerging economies is statistically insignificant in all models.
- The unemployment rate is significant in explaining copper, iron, lead, tin, silver and zinc prices.

Goodness of Fit

		Pure Trend Model	Partial Linear Model		
		$f(t)$	Total	$f(t)$	$z_t\beta$
Oil	1901-2014	0.950	0.960	0.774	0.488
Natural Gas	1992-2014	0.910	0.940	0.753	0.112
Coal	1946-2014	0.960	0.980	0.497	0.640
Copper	1901-2014	0.770	0.810	0.640	0.357
Nickel	1901-2014	0.830	0.830	0.824	0.012
Zinc	1901-2014	0.230	0.260	0.177	0.252
Iron	1901-2014	0.920	0.930	0.867	0.403
Tin	1901-2014	0.800	0.830	0.748	0.066
Silver	1901-2014	0.900	0.910	0.870	0.029
Lead	1901-2014	0.420	0.490	0.356	0.079
Aluminium	1901-2014	0.930	0.930	0.922	0.006

Relationship to the Literatures

- Deterministic trend models
- Super-cycles, spectral analysis
- Stochastic trend models

Concluding Comments

- Trends are functions of unknown shape – hence nonparametric techniques are useful.
- Estimation of trend function can be guided by the ability of the estimator to predict out of sample (cross-validation).
- Commodity trends may display shape similarity, at least over certain periods of time. This can improve precision of estimation and anticipation of certain scenarios.

Hydrocarbon Shape Similarity

- Consider the hydrocarbon triad. From 1946 to 2008 the correlations of the (log) price of oil with gas and coal are 97% and 98% respectively.
- Then gas and coal prices turn down.
- Perhaps the recent drop in oil prices is not entirely surprising.