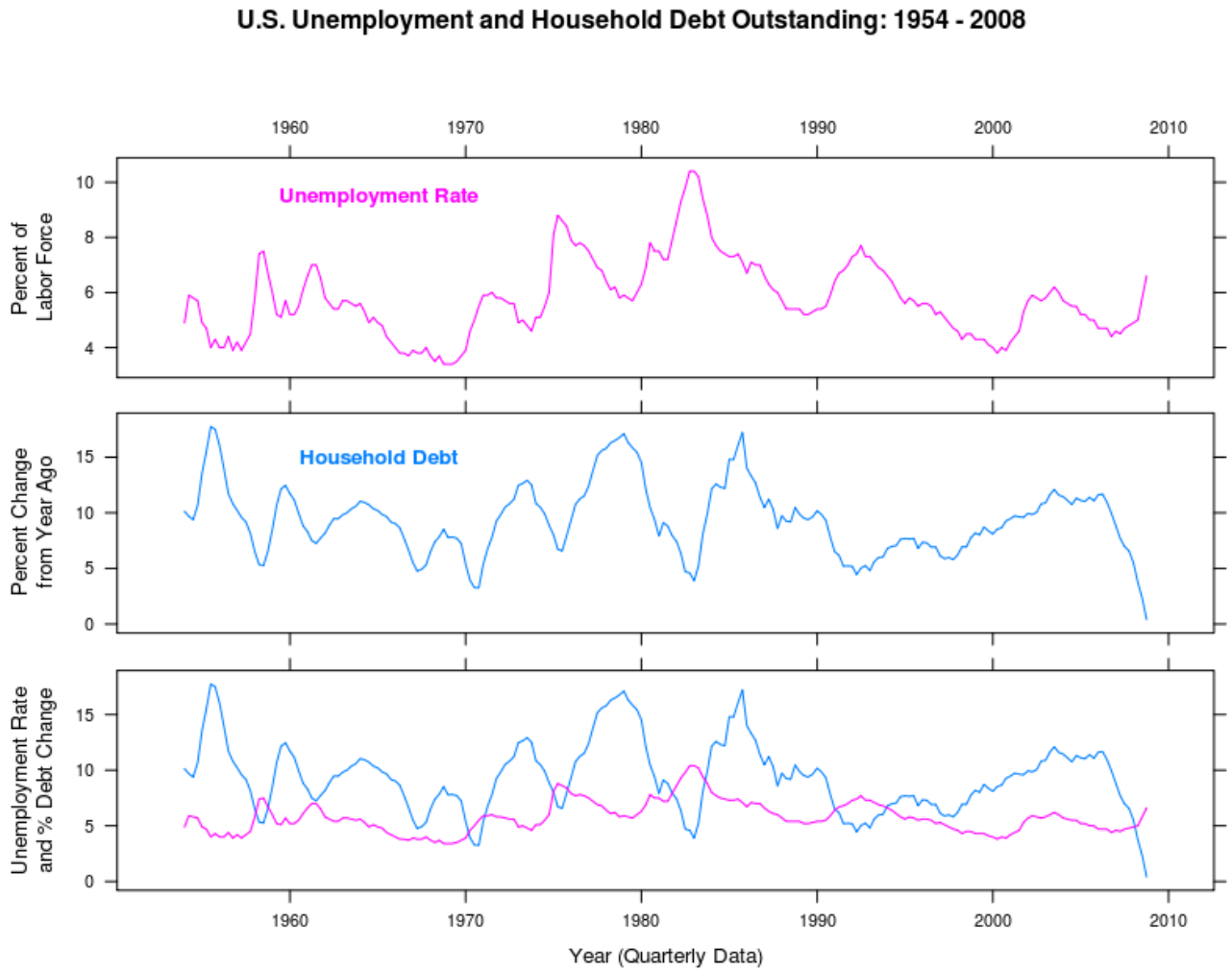
End Result (Sample Analysis)



Data Sources:
Board of Governors of the Federal Reserve System. Series CMDEBT: Household Credit Market Debt Outstanding.
U.S. Department of Labor, Bureau of Labor Statistics. Series UNRATE: Civilian Unemployment Rate.
Both time series were accessed with the API of the Federal Reserve Economic Data (FRED) at the Federal Reserve Bank of St. Louis.
Note: Data was accessed 2009-05-22.

Federal Reserve Economic Data (FRED)

The Economic Research Division of the Federal Reserve Bank of St. Louis recently made their Federal Reserve Economic Data (FRED) database available as XML via an API that uses a REST web service architecture.

Details about the FRED API are available at: http://api.stlouisfed.org/docs/fred/

The FRED database contains data on more than 20,000 economic time series. See details at: http://research.stlouisfed.org/fred2/

For purposes of what follows, here is a brief orientation to FRED:

1. It's available via 20 different HTTP GET requests in 4 areas.
2. The four areas are called Sources (eg, Board of Governors of the Federal Reserve System); Releases (eg, H.8 Assets and Liabilities of Commercial Banks in the United States); Series (eg, Bank Credit of All Commercial Banks); and Categories (eg, Commercial Credit).
3. At the time I explored FRED with XQuery, there were 24 sources, 88 releases, 20,056 data series, and 5014 categories. It's the series that contain the actual data values.
4. 1-to-1 and 1-to-many relationships: Each source may have 1 or more releases. A release usually has just one source, but in a few cases may be a joint release from two or more sources. Each release may have 1 or more times series, but each times series is associated with only a single release. Categories can be nested up to 5 levels deep. A single series may appear in more than 1 category. Note: These statements are my summary; they are not official FRED statements.
5. All sources are available through a single HTTP GET request. The same is true for all releases. However, basic information (eg, id and name) about categories and series is not available in a single GET request. Basic information about all categories and all series must be constructed via the other 20 GET requests.
6. Several of the 20 HTTP GET requests have limits on the number of responses returned in any one request. Offsets can be used to iteratively work through all responses in several GET requests. This is common practice, so no complaints. But it did catch me as I tried to construct basic information about all series using the fred/release/series GET. There are several releases with more than 1000 series and 1000 is the limit for this GET.
7. The FRED web site provides a very nice drill-down to locate series by categories. This offers one useful orientation but little flexibility to create alternative overviews.
8. In order to use the FRED database via the API, you'll need to create a user account with the Federal Reserve Bank of St. Louis and obtain an API key. For details, see: http://api.stlouisfed.org/api_key.html.

XQuery

XQuery is a W3C specification that allows data in XML documents to be queried. A useful resource on XQuery, including requirements for the next version and possible extensions, is available at: http://www.w3.org/XML/Query/

XQuery specifications have been implemented in a number of commercial and open source applications. One of these is Zorba. See details at: http://www.zorba-xquery.com/index.php/about/. Zorba XQuery has several features that I found attractive. One is REST functionality.

Here are some comments about Zorba XQuery based on this project.

1. For software that is not yet at the 1.0 version release, it performed flawlessly and without bugs. The REST calls to the FRED API worked without incident.
2. The learning curve is steep for someone like me coming out of a SQL background and relational databases.

This is not meant to be a criticism. It's just a statement of fact that would be true for any implementation of XQuery. I'd guess I've now written 500 or 600 XQuery programs, but this is easily an order of magnitude too low to achieve skill and comfort with the query tool.

3. I did not conduct rigorous performance tests with XQuery, but query completion times were not impressive. I say this relative to the performance I'd expect with SQL and relational databases on the same machine. Of course, it's not an apples-to-apples comparison to judge XQuery processing local XML files with SQL queries in a tuned database. But, still, performance was sluggish. For example, it took about 195 seconds to complete the FRED overview query (see below). This is a nested query and therefore a little complex, but the size of the XML files is trivial (only 5000 or so nodes in a document accessed multiple times). Again, a disclaimer. I used an old clunker for a server; we're definitely not talking about a fast machine. CPU was the limiting factor. Re-writing queries improved performance in some cases (eg, using full XPaths). But clearly I need to play with Zorba in a number of different situations to better assess performance. These situations might include: a different server; use in an XML database; use in the cloud with Amazon EC2 and S3; and perhaps running XQuery in a browser.

Project

In this project I set out to explore the FRED database and to learn about Zorba's XQuery in the process. There were 2 end products: 1) the capability to generate a variety of overviews of the 20,000 economic times series; and 2) a specific illustration of the use of two of these times series.

1. Overview of FRED Data Series

Below is a screen shot for one of the overviews I created of FRED. It shows the hierarchical classification of time series into categories. In the situation shown, the hierarchy only includes 3 levels. Banking is a level 1 category (ie, its parent is the root category). There are no time series associated directly with Banking. The 8th District Banking Performance is one of the children of the Banking category. It also has no direct time series. However, there are several categories at level 3 that do contain time series. For example, the category Illinois has data in 14 time series.

File Edit View Terminal Help

```
<?xml version="1.0" encoding="UTF-8"?>
<mfred_overview>
  <gen1 name="Banking" id="23" gen="1" num_series="0">
    <gen2 name="8th District Banking Performance" id="64" gen="2" num_series="0">
      <gen3 name="8th District" id="72" gen="3" num_series="7" />
      <gen3 name="Arkansas" id="65" gen="3" num_series="14" />
      <gen3 name="Illinois" id="66" gen="3" num_series="14" />
      <gen3 name="Indiana" id="67" gen="3" num_series="14" />
      <gen3 name="Kentucky" id="68" gen="3" num_series="14" />
      <gen3 name="Mississippi" id="69" gen="3" num_series="14" />
      <gen3 name="Missouri" id="70" gen="3" num_series="14" />
      <gen3 name="Tennessee" id="71" gen="3" num_series="14" />
    </gen2>
    <gen2 name="Commercial Credit" id="101" gen="2" num_series="31" />
    <gen2 name="Condition of Banks" id="83" gen="2" num_series="0">
      <gen3 name="East North Central Census Division" id="86" gen="3" num_series="18" />
      <gen3 name="East South Central Census Division" id="89" gen="3" num_series="18" />
      <gen3 name="Middle Atlantic Census Division" id="85" gen="3" num_series="18" />
      <gen3 name="Mountain Census Division" id="91" gen="3" num_series="18" />
      <gen3 name="New England Census Division" id="84" gen="3" num_series="18" />
      <gen3 name="Pacific Census Division" id="92" gen="3" num_series="18" />
      <gen3 name="South Atlantic Census Division" id="88" gen="3" num_series="18" />
      <gen3 name="United States" id="93" gen="3" num_series="105" />
"mfred_overview.xml" [noeol] 5314L, 398204C                                  1,1           Top
```

The next obvious question is: what time series are available in a given category? Sample results for the Commercial Credit category are shown below.

File Edit View Terminal Help

```
<?xml version="1.0" encoding="UTF-8"?>
<mfred_overview_1>
  <category name="Commercial Credit" id="101" gen="2" parent_id="23" parent_name="Banking">
    <series title="Bank Credit of All Commercial Banks" id="TOTBKCR" observations="1895" release_id=
"22" release_name="H.8 Assets and Liabilities of Commercial Banks in the United States">
      <source name="Board of Governors of the Federal Reserve System" id="1" />
    </series>
    <series title="Consumer Revolving Credit Owned by Commercial Banks" id="REVOLNCB" observations="
495" release_id="14" release_name="G.19 Consumer Credit">
      <source name="Board of Governors of the Federal Reserve System" id="1" />
    </series>
    <series title="Consumer Revolving Credit Owned by Credit Unions" id="REVOLNCU" observations="303
" release_id="14" release_name="G.19 Consumer Credit">
      <source name="Board of Governors of the Federal Reserve System" id="1" />
    </series>
    <series title="Consumer Revolving Credit Owned by Finance Companies" id="REVOLNFC" observations=
"292" release_id="14" release_name="G.19 Consumer Credit">
      <source name="Board of Governors of the Federal Reserve System" id="1" />
    </series>
    <series title="Consumer Revolving Credit Owned by Nonfinancial Businesses" id="REVOLNNFC" observ
ations="471" release_id="14" release_name="G.19 Consumer Credit">
      <source name="Board of Governors of the Federal Reserve System" id="1" />
    </series>
"t.out" [noeol] 98L, 7924C                                                    1,1           Top
```

XQuery Used to Produce the FRED Overviews

Note: The first xquery listed in each step generates an XML document as output. The second xquery checks for obvious inconsistencies that signal an error. The source code is available through the links below. Note that any source code that accesses the FRED API will require your own API key.

Step 1: Retrieve all sources
mfred_sources.xq & count_mfred_sources.xq

Step 2: Retrieve all releases
mfred_releases.xq & count_mfred_releases.xq

Step3: Retrieve source information for all releases
mfred_releases_sources.xq & count_mfred_releases_sources.xq

Step 4: Retrieve all categories
a. mfred_root_category.xq (retrieves root category of use in step 4b)
b. mfred_categories.xq & count_mfred_categories.xq

Step 5: Retrieve series information for all categories
mfred_categories_series.xq & count_mfred_categories_series.xq

Step 6: Retrieve all series
mfred_series.xq & count_mfred_series.xq

Step 7: Retrieve all releases for all series
mfred_series_releases.xq & count_mfred_series_releases.xq

Step 8: Retrieve counts of data observations for all series
mfred_series_observations_count.xq & count_mfred_series_observations_count.xq

Step 9: Augment all series with information on releases and observations counts
mfred_series2.xq & count_mfred_series2.xq

Step 10: Retrieve overview information
a. mfred_overview.xq to generate the first overview shown above.
b. mfred_overview_1.xq to generate the second overview shown above.

2. Sample Analysis using FRED, XQuery and R

Step 1: Retrieve data values for specific time series
a. mfred_1series_observations.xq for UNRATE retrieves data on unemployment rates in the United States. The times series is monthly, with data from 01-January-1948 to 01-April-2009 when I accessed the data on 22-May-2009. The source for this data is the Bureau of Labor Statistics in the U.S. Department of Labor. The output of the query is an XML file.
b. mfred_1series_observations.xq for CMDEBT retrieves data on Household Credit Market Debt Outstanding in the United States. This series is quarterly, with data from 01-January-1953 to 01-October-2008. The source for this data is the Board of Governors of the Federal Reserve System. Again, the output is an XML file.

Step 2: Perform calculations and combine time series
a. mfred_CMDEBT_pctchg.xq calculates a percentage change in the Household Credit Market Debt Outstanding compared to one year ago (eg, from 01-July-2007 to 01-July-2008). Data is quarterly and output is an XML file.
b. mfred_CMDEBT_pctchg_UNRATE.xq combines unemployment and household debt data into a single file covering 01-January-1954 to 01-October-2008. Data for both time series are quarterly at this point. The output is an XML file that can be parsed and analyzed using the R statistics software.

Step 3: Conduct analysis with R
The results from the R analysis are shown in the opening screen shot on page 1. The interpretation I'll leave to others with more knowledge about labor markets and financial data. But it appears that the current recession is unlike any in the past 55 years. The rate of increase in household credit market debt outstanding has slowed during earlier periods when unemployment rose sharply, but never before has the rate of increase screeched to a near halt as it has recently.